

AD-A096 848

CONTROL DATA CORP MONTEREY CA
SURVEY OF AVAILABLE SYSTEMS FOR
FEB 81 P A HARR, W D GROSSUP

IDENTIFYING SYSTEMATIC ERRORS I--ETC(U)
N00228-79-C-X554

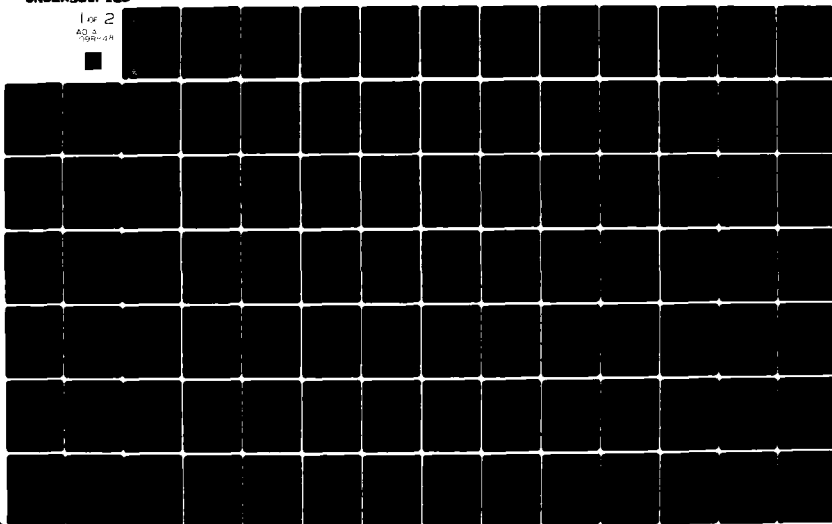
F/G 12/1

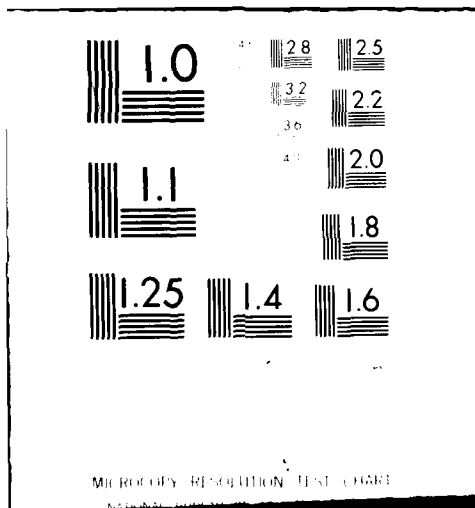
NL

UNCLASSIFIED

NEPRF-CR-81-01

1 of 2
AD-A096 848







NAVENVPREDRSCHFAC CR 81-01

AD A668848

NAVENVPREDRSCHFAC
CONTRACTOR REPORT
CR 81-01

LEVEL II

(12)

SURVEY OF AVAILABLE SYSTEMS FOR IDENTIFYING SYSTEMATIC ERRORS IN NUMERICAL MODEL WEATHER FORECASTS

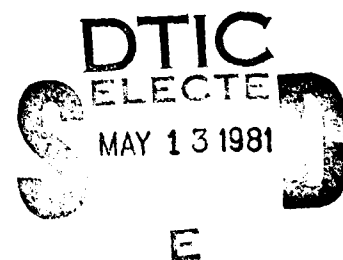
Prepared By:

Patrick A. Harr and William D. Groscup

Control Data Corporation
Monterey, California 93940

Contract No. N00228-79-C-X554

FEBRUARY 1981



DTIC FILE COPY



APPROVED FOR PUBLIC RELEASE
DISTRIBUTION UNLIMITED

81 5 13 006

Prepared For:

**NAVAL ENVIRONMENTAL PREDICTION RESEARCH FACILITY
MONTEREY, CALIFORNIA 93940**

QUALIFIED REQUESTORS MAY OBTAIN ADDITIONAL COPIES
FROM THE DEFENSE TECHNICAL INFORMATION CENTER.
ALL OTHERS SHOULD APPLY TO THE NATIONAL TECHNICAL
INFORMATION SERVICE.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER NAVENVPREDRSCHFAC Contractor Report CR 81-01	2. GOVT ACCESSION NO. AD-A098	3. RECIPIENT'S CATALOG NUMBER 848
4. TITLE (and Subtitle) Survey of Available Systems for Identifying Systematic Errors in Numerical Model Weather Forecasts		5. TYPE OF REPORT & PERIOD COVERED 9 Final
6. PERFORMING ORG. REPORT NUMBER CR 81-01		7. AUTHOR(s) 10 Patrick A. Harr and William D. Groscup
8. CONTRACT OR GRANT NUMBER(s) 1. N00228-79-C-X554		9. PERFORMING ORGANIZATION NAME AND ADDRESS Control Data Corporation Monterey, CA 93940
10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS PE 63207N, PN TW0513 TA CC00, NEPRF WU 6.3-13		11. CONTROLLING OFFICE NAME AND ADDRESS Naval Air Systems Command Department of the Navy Washington, DC 20361
12. REPORT DATE February 1981		13. NUMBER OF PAGES 94
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) Naval Environmental Prediction Research Facility Monterey, CA 93940		15. SECURITY CLASS. (of this report) UNCLASSIFIED
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES Original manuscript received April 1980.		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Systematic errors Numerical forecast models Model verification		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) A generalized and flexible software plan to identify systematic errors in numerical forecast models is described. The plan is applicable to different types of models and can be altered to operate in different modes. Examined are various techniques which can be used within the plan, in terms of advantages and disadvantages. Estimates are made of possible impacts upon an existing operational environment and of the usage of resources needed by a system for identifying systematic errors.		

DD FORM 1473

EDITION OF 1 NOV 65 IS OBSOLETE
S. N. 0102-014-6A01

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

TABLE OF CONTENTS

	<u>Page</u>
SECTION 1.0 INTRODUCTION	1
1.1 Objectives	1
SECTION 2.0 A MODEL VERIFICATION SYSTEM	4
2.1 The General Verification System	5
2.2 Basic Analysis	9
2.3 Detailed Analysis	11
2.4 Software Details	12
2.4.1 Input Processing	15
2.4.2 Analysis	16
2.4.3 Output Processing	16
SECTION 3.0 TECHNIQUES APPLICABLE FOR MODEL VERIFICATION	22
3.1 Basic Techniques	22
3.1.1 General Measures	22
3.1.1.1 Mean Square Error and Standard Deviation	23
3.1.1.2 Correlation	27
3.1.1.3 Brier Score	28
3.1.2 Graphical Methods	29
3.1.2.1 Difference Fields	29
3.1.2.2 Longitude-time plots (Hovmöller diagrams)	30
3.1.2.3 Zonal averages and Meridional cross sections	32
3.1.3 Other Specific Measures	32
3.1.3.1 S1 Score	38
3.1.3.2 Threat Score	38
3.1.3.3 Prefigurance, Post Agreement and Heidke Skill Score	41
3.1.4 Summary of Basic Techniques	42
3.2 Detailed Techniques	42
3.2.1 Spectral Analysis	42

TABLE OF CONTENTS

CONTINUED

3.2.2	Quasi Lagrangian Analysis	46
3.2.3	Quasi Geostrophic Analysis	47
3.2.4	Regression Techniques	47
3.2.5	Pattern Recognition	48
3.2.6	Diagnostic Studies	54
3.3	Specific Techniques	57
3.4	Summary	60
3.4.1	Advantages, Disadvantages of techniques	60
3.4.1.1	Basic Techniques	61
3.4.1.2	Detailed Techniques	65
3.4.2	Conclusions	69
SECTION 4.0	SYNOPTIC SCALE VARIABLE USED FOR MODEL VERIFICATION	70
4.1	Basic Analysis	70
4.2	Detailed Analysis	71
4.3	Summary	72
SECTION 5.0	COST ESTIMATION	73
5.1	Impact Summary	73
5.1.1	Equipment Impacts	73
5.1.2	Software Impacts	74
5.1.3	Organizational Impacts	74
5.1.4	Operational Impacts	75
5.1.5	Developmental Impacts	75
5.1.6	Conclusion	76
5.2	Computer Resources	76
5.2.1	Basic Analysis Resources	76
5.2.2	Detailed Analysis Resources	78
5.3	Development Resources	79
5.4	Cost Summary	80

TABLE OF CONTENTS

CONTINUED

SECTION 6.0	CONCLUSIONS AND SUGGESTIONS FOR FURTHER STUDY	81
6.1	Summary	83
REFERENCES		87
Appendix A		90

Accession For		
NTIS GRA&I	<input checked="" type="checkbox"/>	
DTIC TAB	<input type="checkbox"/>	
Unannounced	<input type="checkbox"/>	
Justification		
Distribution/		
Availability Codes		
and/or		
Special		
A		

1.0

INTRODUCTION

Numerical models now provide much of the foundation for the synoptic scale predictions from which the United States Navy derives a large group of environmental products. Due to the complexity of the techniques and methodology of these models and the atmosphere they attempt to simulate, it has become increasingly difficult to objectively determine the relative value of a model's simulation. Operational priorities and the stochastic nature of the atmosphere can dictate that any improvement in a model performance be only partially dependent upon the physical nature of the system. This fact makes the improvement of verification statistics a basic criterion for model improvement which can provide insight into the accuracy of forecast parameters. Knowledge of any forecast errors or biases will also provide several additional benefits such as: methods of comparison of one model against another; discovery of logic or code errors in existing models plus validity of simplifying assumptions in models.

Another problem is the fact that errors which occur in a forecast system, or any type of system for that matter, can be systematic (ie. forced) or random. We define a systematic error as an error which can be detected and examined in relation to possible sources as opposed to random errors which cannot be determined in relation to possible sources.

The purpose of this report is to document a system which can be used to examine systematic errors in numerical model forecasts.

1.1

Objectives

The objectives of this study are the following:

- i) To describe a software flow which describes the procedures that could be used to identify systematic errors in a numerical model;

- ii) Examine and describe various techniques which could be used to identify systematic errors in a numerical model;
- iii) Describe what model parameters should be used to help identify errors in the model forecasts;
- iv) Estimate a cost, in terms of impacts and resource usage of a software system which identifies systematic errors in a numerical model forecast.

Section two contains a step by step plan which outlines the procedures, within the verifying system. Section three contains descriptions of various techniques which can be used within the verifying system. Section four describes the variables which the system can use in order to examine the numerical product. Section five outlines the cost of the verifying plan. The last section contains recommendations for further study and implementation of a verifying system.

Appendix A contains brief descriptions of a number of numerical models currently in use at a number of large weather facilities and used in this report to illustrate the application of various techniques described within the text. Models used for illustrative purposes are used at the following meteorological facilities:

- i) National Center for Atmospheric Research (NCAR)
- ii) Geophysical Fluid Dynamics Laboratory (GFDL)
- iii) National Meteorological Center (NMC)
- iv) Goddard Institute for Space Studies (GISS)
- v) University of California at Los Angeles (UCLA)
- vi) Canadian Meteorological Center (CMC)

A large number of verification studies have been applied to these models and are documented in the meteorological literature. Examinations of these types of studies and how they have utilized various methods which are capable of identifying systematic errors in numerical forecasts can provide insight

into the merit of various techniques. Where deemed necessary examples of techniques, used in these studies, are presented within this report in order to provide more insight into the use of the specific technique.

It is important before any verification system is designed that specific guidelines be set up as to the purpose of any component of that system (eg. statistical technique or model parameter). Because there are many types of statistical techniques which can be applied to many data sets or combination of data sets, a strict set of guidelines will prevent any over/under analysis of a specific parameter or the inclusion of many techniques which do not perform the required or anticipated analysis. For example, Tracton and Stackpole (1976) have described the guidelines of the National Meteorological Center's (NMC) verification program as follows:

- i) To identify and diagnose critical problems in the operational analysis and forecast systems;
- ii) Provide overall objective measures of the absolute and relative skill of analysis and forecast systems;
- iii) Document the performance characteristics of analysis and forecast systems.

These guidelines are quite general and should apply to any verification system at a large environmental center.

Two additional guidelines of a verifying system designed for use at Fleet Numerical Oceanography Center (FNOC) are that the system concentrate on the synoptic scale features of the numerical products and that the system be user oriented. The constraint of concentrating on synoptic scale features is straight forward and does not require any further comment here.

Regarding the latter constraint of being user oriented we can divide the FNOC product users into the following groups:

- i) Users who are primarily interested in a particular geographical region;

- ii) Users who are mainly interested in a global (hemispheric) situation;
- iii) Users primarily involved in aviation;
- iv) Users who are primarily ocean going;
- v) Users who are constrained to a rigid operational time schedule;
- vi) Users not restricted to a rigid time schedule.

Therefore the ideal system would be one which could supply to each user group listed above the information required most by each respective group. Of course it may be impossible to supply this information to each group for each time, however in order to maintain a maximum of utility and sophistication we suggest that a verification system have as a minimum the following characteristics:

- i) Timeliness;
- ii) Lends itself to rigorous interpretation;
- iii) Can identify critical problem areas;
- iv) Can document specific performance characteristics;
- v) Can diagnose the product in terms of the physical, dynamical and computational aspects;
- vi) Can be spectral;
- vii) Objective;
- viii) Statistical;
- ix) Plausible;
- x) Regional as well as global.

The incorporation of these attitudes would provide a comprehensive evaluation system which would be user oriented.

2.1 The general verification system

A two staged verification system is best suited to incorporate the ten characteristics mentioned above, plus provide the versatility to meet the needs of the wide variety of FNOC users. Figure 2.1 provides a schematic outline of a

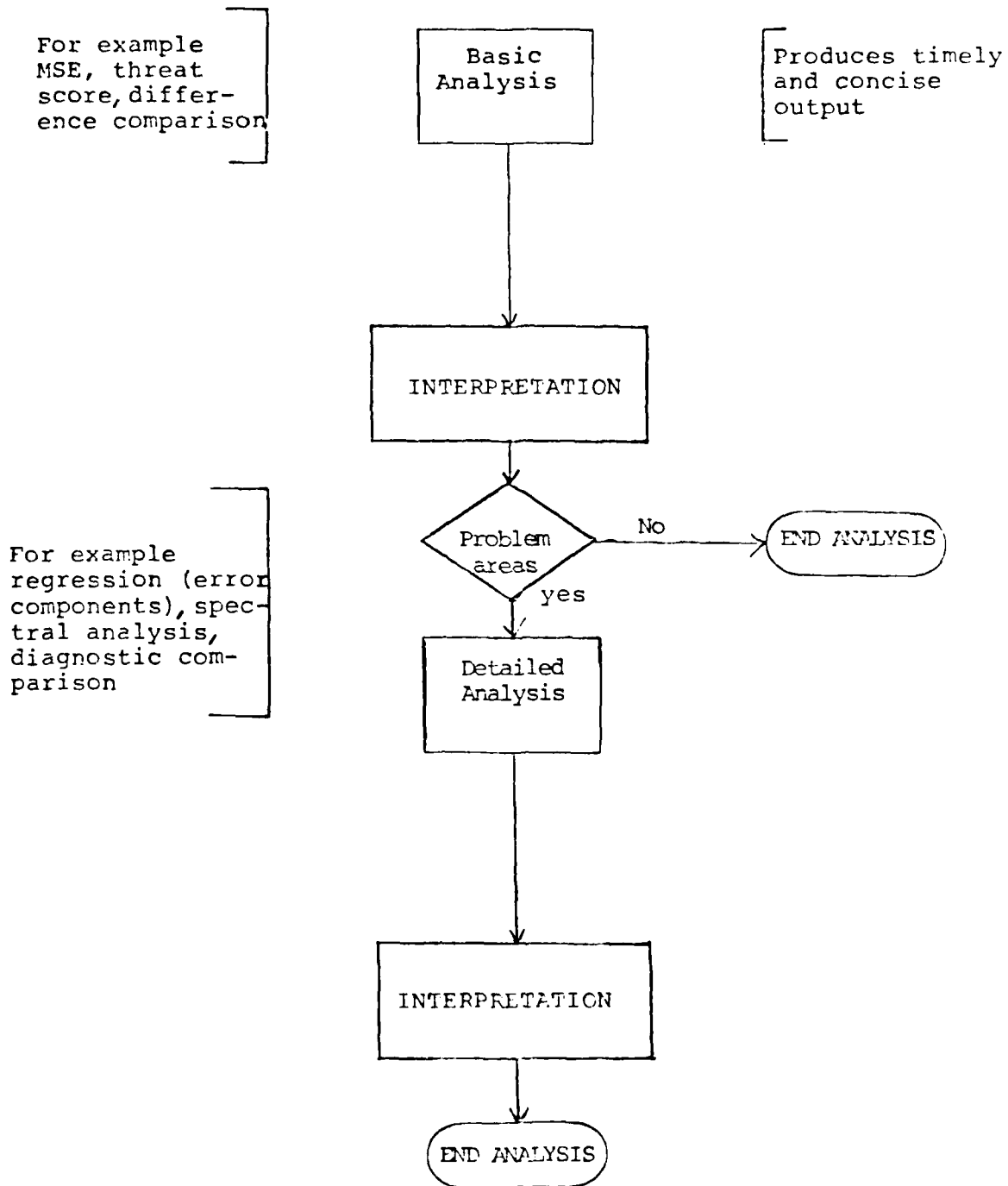


Figure 2.1
A two staged verification system

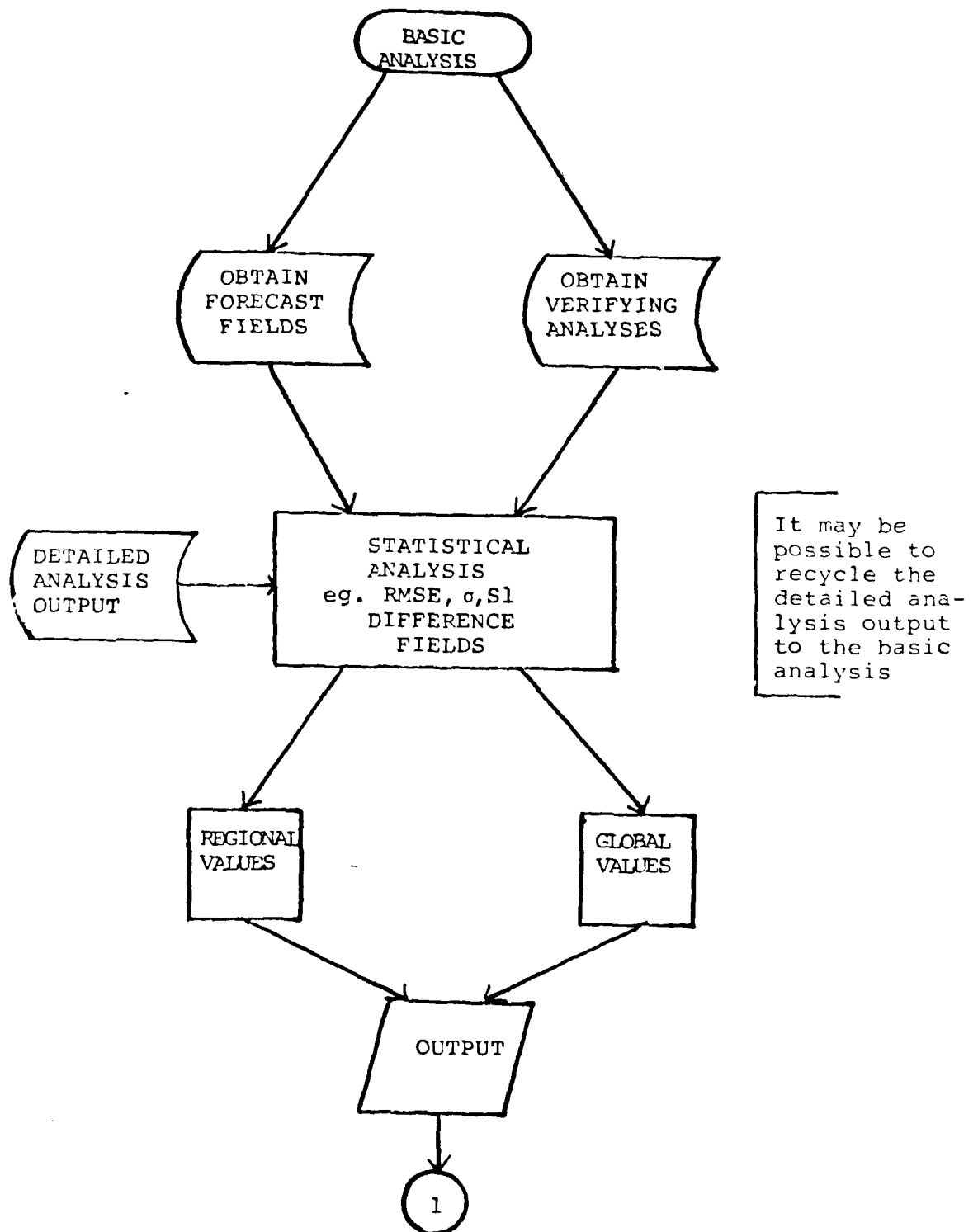


Figure 2.2

Schematic diagram for the Basic Analyses

distribution
of output
to product
users

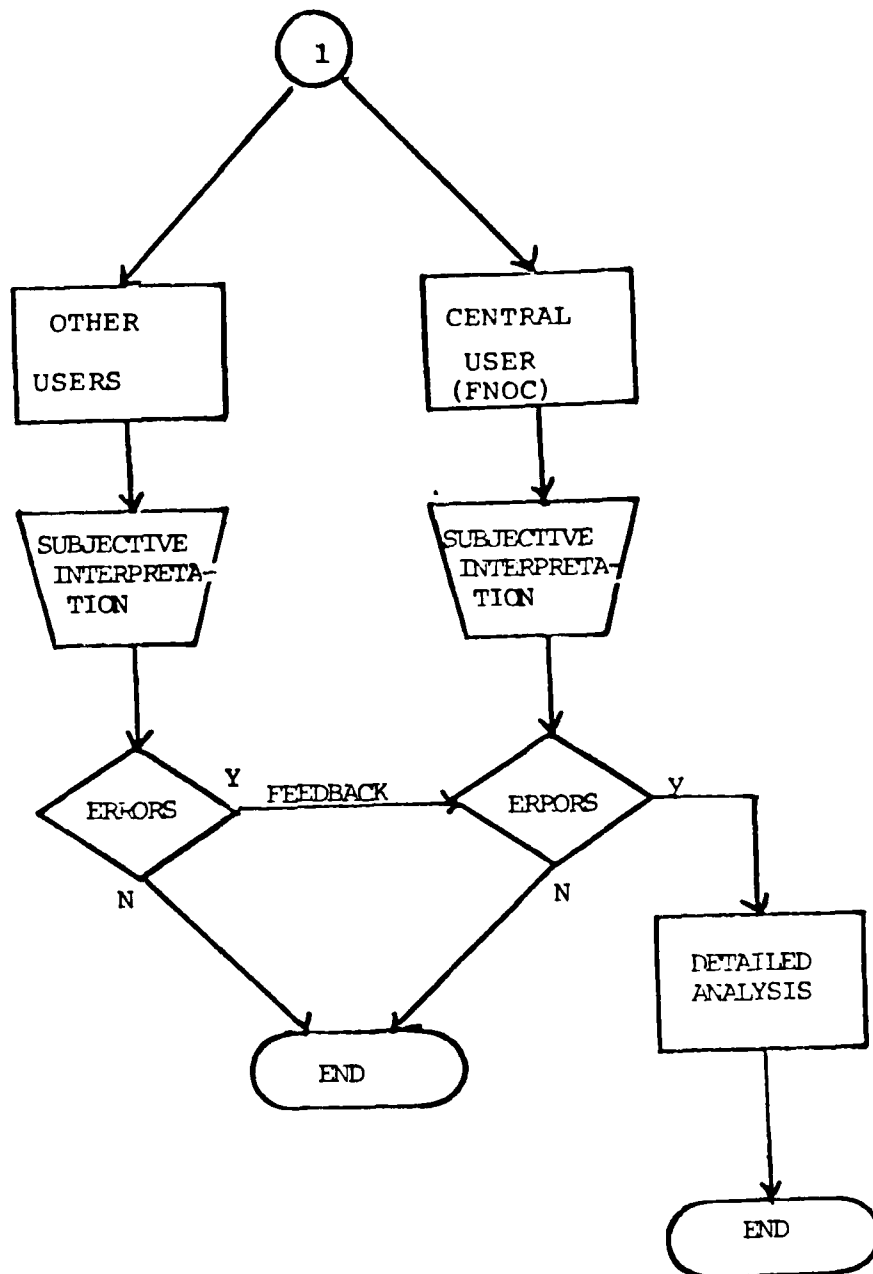


Figure 2.2
(continued)

two staged verification system.

As the figure shows, this type of system is designed to meet the following generalized objectives;

- i) Perform a basic analysis, providing an output which is in a form which can be easily interpreted and diagnosed in terms of error locations and magnitudes;
- ii) Capability of providing a more sophisticated analysis if deemed necessary. This analysis will allow an assessment of model strengths and weaknesses in areas pointed out by the basic analysis, to be in error.

2.2 Basic Analysis

The basic analysis would provide a relatively quick look at the model performance. This would be accomplished through the use of simple statistical techniques, or measures and graphical interpretations.

The most important component of the basic analysis is the specification that the analysis provide a timely and concise output which is easily interpreted and can accommodate users who are under an operational time constraint. This would highlight interesting features or recurring errors which could then be examined in more detail if deemed necessary.

A schematic diagram of the flow of the basic analysis is presented in figure 2.2. Obtaining the appropriate data can be accomplished through the use of the standard FNOC software components. Details of this process are described in section 2.3. Timely, concise and easily interpreted output will be distributed to users for the purpose of improving their forecasts and environmental products. Subjective interpretation will then be performed by users who have an interest in certain parameters or a specific geographical region. FNOC will be responsible for interpretation of the entire analysis. Respective users shall interact with FNOC in terms of feedback on any problem areas or unusual results. FNOC

will also determine if any problem areas are present, in terms of recurring error patterns or other features. FNOC will then decide if a detailed analysis is needed to further examine certain features or parameters. This may involve specifying that the data be saved for a future analysis.

It is evident from the above discussion that the timeliness of this process, necessary for an improvement in the environmental products, is dependent upon the interpretation of the basic analysis result. This stresses the importance of a concise and straight forward output package including plotted and printed results.

Initially, the determination of problem areas or recurring error patterns necessary for a decision to apply a detailed analysis may be better left up to the central organization (FNOC). However as other users receive analysis results and direct their feedback to FNOC the evaluation process may involve more users. This is necessary because certain users may be more experienced in the meteorological conditions in their specific geographic area and should be able to supply FNOC with a good interpretation of the verification statistics.

In summary the basic analysis would involve the following procedure:

- i) Obtain forecast and verification data;
- ii) Perform specified statistical evaluations and manipulations;
- iii) Process timely and concise output to respective users;
- iv) Enable interpretation by respective users on their respective areas and an overall interpretation by FNOC;
- v) Enable a decision as to saving data and/or running a more detailed analysis, based on FNOC evaluation or feedback from users.

2.3 Detailed Analysis

The detailed analysis will provide an evaluation of a model's higher order terms which can be derived from the more basic variables. This analysis would provide greater insight into the model's performance than offered in the basic analysis and would be of more use to, perhaps, a model research and development group.

The detailed analysis will be implemented after the basic analysis is completed. The time lag between the two analyses would depend upon a number of parameters such as data availability and severity of the error. For instance, if a persistent problem has been detected in the basic analysis, but it is not interpreted as severe, the detailed analysis performed on a specific data set could be run at some future time. However, a ceiling should be placed on the time between the basic analysis and when the detailed analysis is performed. This prevents the excess build up of unused data. Of course, if a severe error is detected the detailed analysis could be run as soon as possible.

A schematic diagram of the flow of the detailed analysis is shown in figure 2.3. Input to this analysis can be organized identically as the input to the basic analysis (see section 2.4). Of course the actual analysis will be quite different. The other major difference between the two analyses will be in the organization of the output. The output of the detailed analysis needs to be as clear and concise as the output for the basic analysis. The detailed output should be organized in a manner that will highlight the specific feature under examination in the particular technique. The output of the detailed analysis should be arranged in a manner that will provide easy interpretation of results upon inspection. Results, output in this form, may be "recycled" into the basic analysis which can organize analysis results in a simply organized format which allows for straight forward interpretation.

It is expected that the detailed analysis results will be examined only by those users who have a more thorough knowledge of the numerical model and meteorological conditions.

In summary the detailed analysis would involve the following procedures;

- i) Input data specified from the basic analysis;
- ii) Evaluation of specific techniques;
- iii) Interpretation of the output by the central user and possibly other users or development group;
- iv) Recommendations or changes which may lead to model improvement and are based on the detailed analysis results.

We have now described the software flow of the two staged verification system. No attempt has been made to describe techniques or measure what would be included within the flow of the system. This will be discussed in section 3. However, it should be noted here, that the inclusion or deletion of techniques to and from the analysis should be flexible.

The flow of the verification analysis will begin with a basic examination of the numerical model forecast employing the use of proven statistical measures. This shall be examined by personnel who have a detailed knowledge of certain meteorological conditions and the meteorology of specific geographical areas. There will also be an examination by personnel who are familiar with the workings of the operational model. These examinations may lead to the performance of a more detailed analysis which will be examined by personnel who are familiar with the details of the numerical model and meteorological conditions. This will lead to improvements in the numerical model and environmental products.

2.4 Software details

We can make the following guidelines for the computer software which would perform as described above.

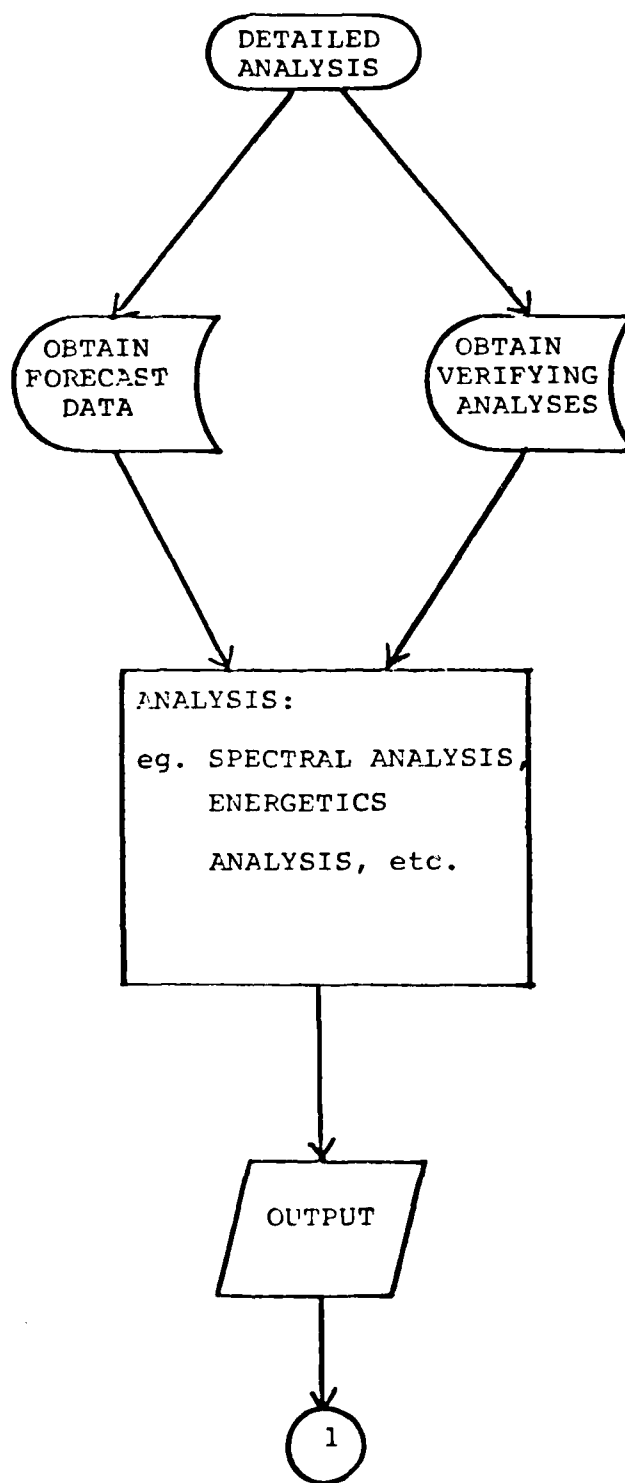


Figure 2.3

Schematic diagram for the detailed analysis

Dashes represent optional flow. Output may be distributed to other users or a possible model development group.

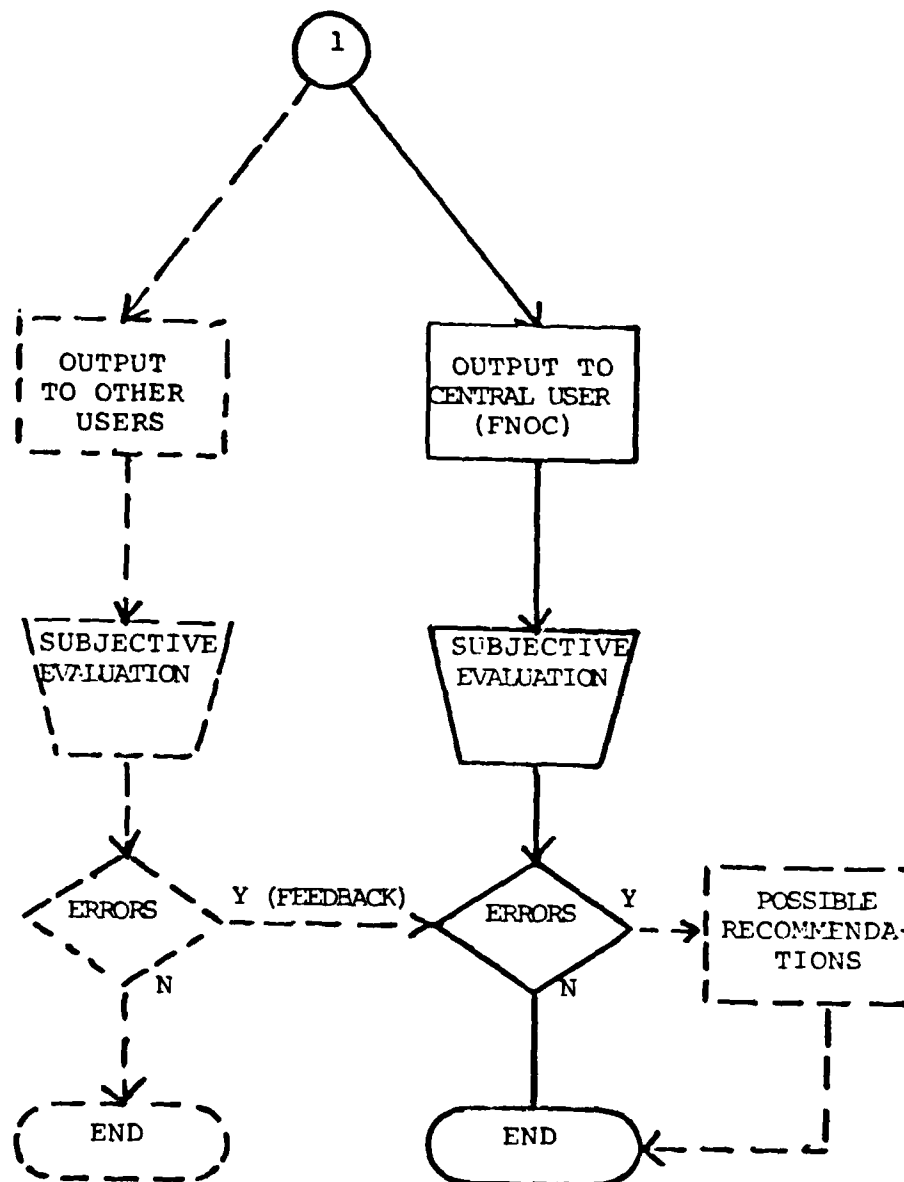


Figure 2.3 (continued)

- The software can be divided into three sections;
- i) Input processing;
 - ii) Analysis;
 - iii) Output processing

2.4.1 Input processing

It would be desirable to provide the personnel in charge, with control over the operation of the verification system. This is accomplished by specifying that directives, such as what data is to be used or what methods of analysis should be employed, can be used to control the running of the verification system. This is necessary in order to avoid processing too much data or unnecessary measures or plots. This will keep the wasting of resources at a minimum.

The input module would basically be the same for both the detailed and basic analyses. This section would access the FNOC input/output software component used for environmental data. The actual data fields required could be variable, depending upon past results or which analysis mode is currently operating (e.g. basic or detailed). The program user should be able to specify which data needs to be obtained through input data to the program.

The input processing for the verification analysis may be designed to sustain a data stack which would be maintained in the following manner. The input software would obtain the current analysis field which would be used to verify the appropriate 24, 48 and 72 hour forecasts which would be revolved into and out of the analysis. For instance, on 10 January the data stack would contain, say, heights of 1000 mb and 500 MB pressure surfaces for:

10 January analysis - comparing to:

- 9 January 24 hr. forecast
- 8 January 48 hr. forecast
- 7 January 72 hr. forecast

On the next day, 11 January the data stack would contain

11 January analysis - comparing to:

- 10 January 24 hr. forecast
- 9 January 48 hr. forecast
- 8 January 72 hr. forecast

It is therefore necessary that the forecast data fields be saved for the amount of time necessary for their verification (e.g. 72 hr. forecast would be saved for at least 72 hours). The user would have the option of requesting forecast and/or analysis fields to be saved for a longer period within the verification data stack. This is necessary to allow the analysis section to display time sequenced events such as Hovmöller diagrams where a number of days of data are required. An idealized flow of the input process is shown in figure 2.4.

2.4.2 Analysis

The analysis section would vary for the basic and detailed analyses. This section shall evaluate the specific techniques for each analysis respectively. The analysis section could be constructed of a number of routines, each evaluating a particular measure or technique. The software should be modular enough to allow deletion and/or additions of methods as determined necessary by the organization performing the verification analysis. Routines would also prepare data for plotted output by arranging data into the proper format for input into the various plotting software components.

A main constraint upon this section would be the ability to regionalize all calculations by input data to the program. For instance, the program user should be able to explicitly determine the geographical domain for specific calculations of the analysis techniques. This is a particularly important point for the basic analysis.

The analysis section shall access an input file, specified by the user responsible for the system operation that will direct the analysis of the data by specifying which techniques should be applied or deleted from the analysis.

2.4.3 Output Processing

The output section will involve two components; printing and plotting of analysis results and distribution of results to specific organizations.

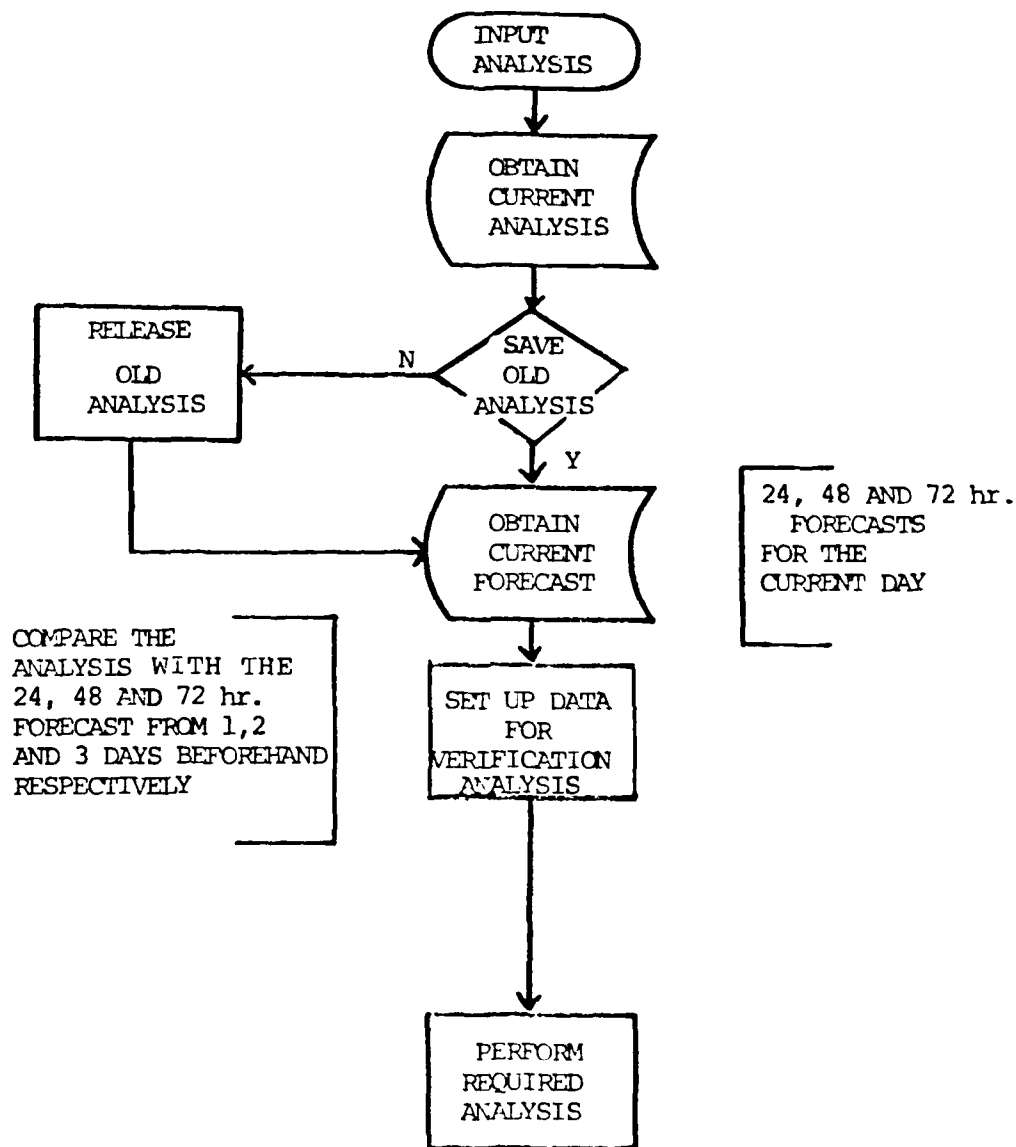


Figure 2.4
Generalized flow chart for the input section

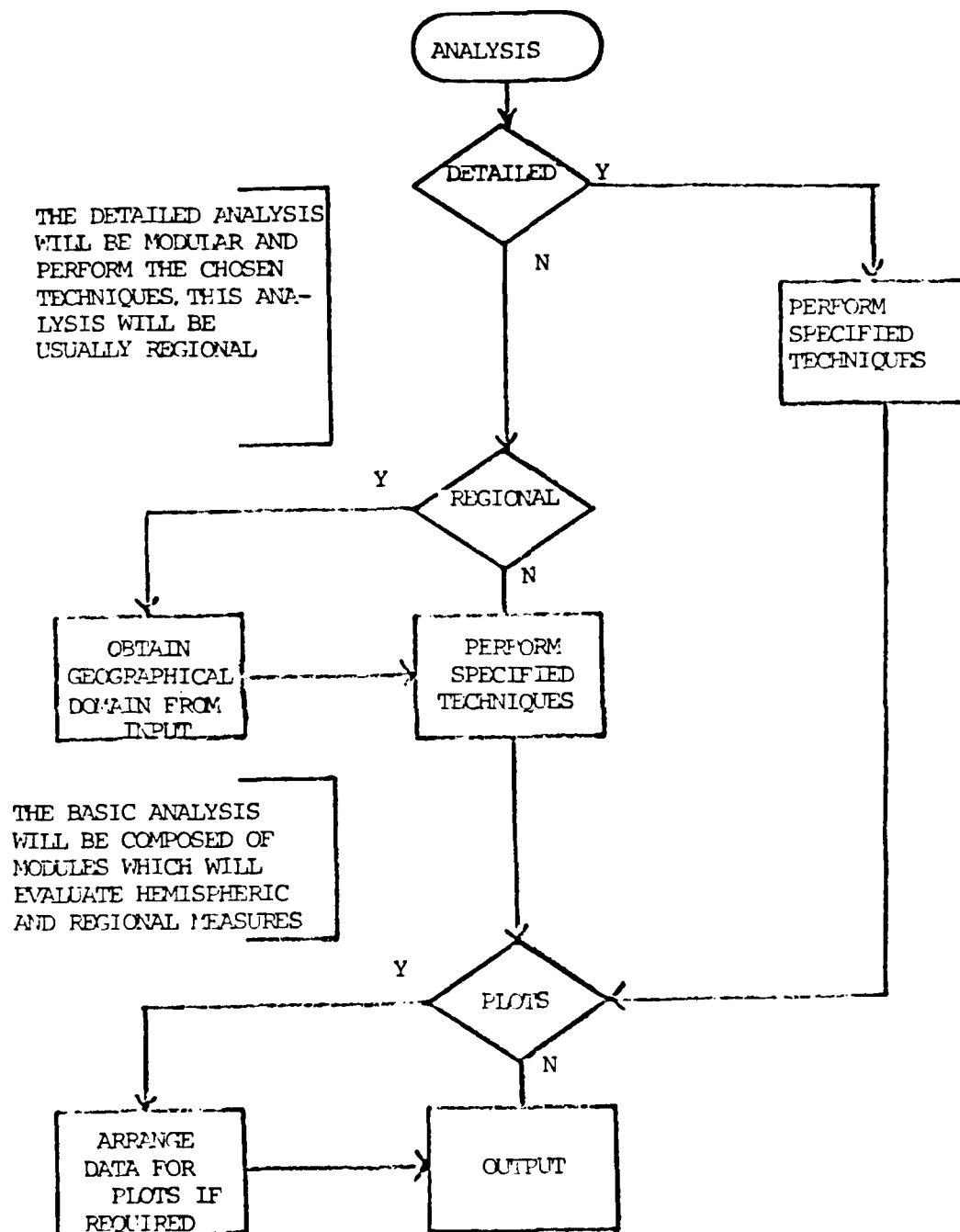


Figure 2.5

Generalized flow chart for the analysis section

The printed output for the basic analysis would be taken from the analysis results as to portray a clear and concise picture of the model performance. The output should be self contained including labels explaining all results printed for a particular technique or model parameter. Plotted output for the basic analysis may consist of contoured difference fields or time/height plots of specific parameters. These figures need to be output in a easily interpreted manner. Much of the software needed to provide plots exist in the VARIMAP package.

The printed and plotted output for the detailed analysis need not be as concise as the basic analysis components. The output for the detailed analysis shall be arranged in a manner which will highlight the particular feature of the specified technique. For example, spectrally analyzed data shall be output in a manner which will allow comparisons of the various wave modes for specific model parameters.

A separate output module shall arrange for the distribution of the basic analysis to other users over the Naval Environmental Data Network. This might require a specialized output format such as width of the output page for example. Existing software packages, available at FNOC, can be used to prepare printed and plotted output over the NEDN system.

A final responsibility of the output section would be to decide if it is necessary to retain the data fields used for the verification run. This can be specified by input directives from the program user. A limit to the number of forecasts that may be saved should be made. Data no longer needed can be added to the climatology data base at FNOC or released if not normally retained.

Figure 2.6 shows a generalized flow chart for the output processing.

All three sections described will work to identify systematic errors in a numerical model forecast and can operate within the FNOC software specifications.

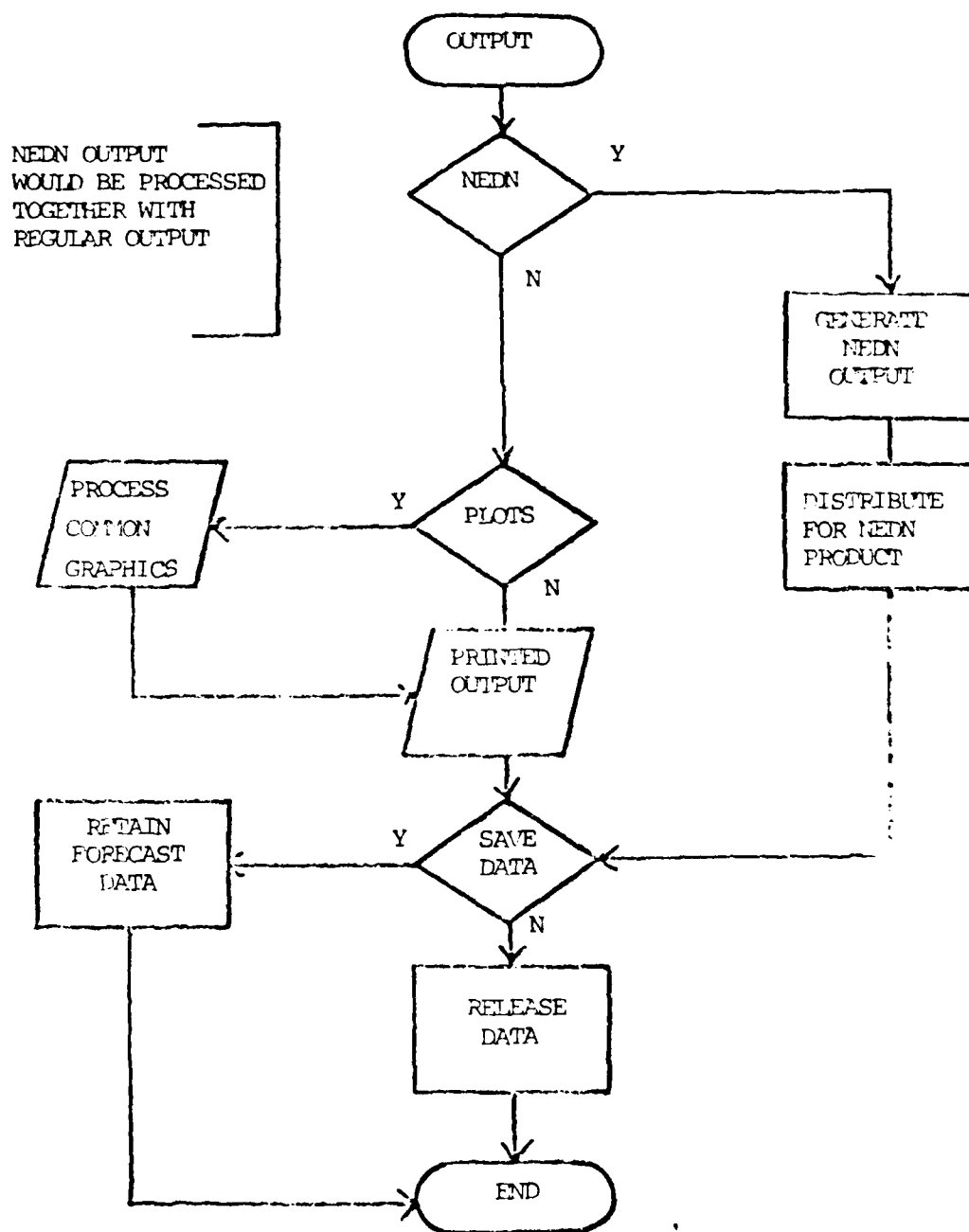


Figure 2.6

Generalized flow chart for output processing

This system must function within an operational framework. This fact stresses the need for the interpretation of the output to be easy and straightforward. Therefore the output module must provide clear and concise output results for even the more complicated detailed analysis results. This is the single most important feature necessary for the success of the verification system.

3.0

TECHNIQUES APPLICABLE FOR MODEL VERIFICATION

In the previous section we divided a verification scheme into a basic and detailed analysis. We can also divide various techniques into the same type of categories. Basic techniques which are primarily statistical and applicable in the basic analysis of Section 2 and more detailed techniques, which are not always statistical but provide diagnostic studies of model simulations can be grouped respectively. Specific advantages, disadvantages and uses of the various techniques are discussed at the end of the section.

3.1

Basic Techniques

There are a number of basic techniques which are applicable to model verification. Most of these are statistical. We can group these techniques into the following categories:

- i) General measures which are based upon widely used statistical parameters (e.g. correlation);
- ii) Graphical interpretations;
- iii) Other general measures which are not always based on statistical parameters (e.g. SI score).

3.1.1

General Measures

A typical simple verification system might start with a measure of the difference between a forecast, F_i , parameter and an observed parameter, O_i as

$$d_i = F_i - O_i$$

This value is then used to define the mean square error (also known as the performance) as

$$MSE = \frac{1}{N} \sum_i d_i^2 ,$$

Where N is the number of points in the field. Consequently we can determine a bias, that is a measure of whether the model is over or under forecasting;

$$\text{Bias} = \bar{D} = \frac{1}{n} \sum d_i.$$

These are all basic measures which have been widely used at various facilities (Daley, 1976; Bengtsson, 1976; Baumhefner, 1976).

3.1.1.1 Mean Square Error and Standard Deviation

A more exact formulation of the MSE is

$$\text{MSE} = \frac{\sum_i \frac{(F_i - O_i)^2}{M_i^2}}{A}$$

Where $A = \sum \frac{1}{M_i^2}$ and M_i is the appropriate map factor.

The map factor needs to be applied when data is stored in a latitude/longitude grid.

The root MSE (RMSE) is defined by taking the square root of the MSE.

This measure is the basis for many verification systems and studies. An illustration of this score is shown in figure 3.1. This figure also demonstrates the utility of the MSE measure in that it can be used for many types of models, including spectral.

The standard deviation, σ , is related to the MSE and is defined as follows;

$$(N) \sigma \text{ (forecast field)} = \left[\sum (F_i - \bar{F}_i)^2 \right]^{1/2}$$

$$(N) \sigma \text{ (analysis field)} = \left[\sum (O_i - \bar{O}_i)^2 \right]^{1/2}$$

$$(N) \sigma \text{ (error field)} = \left[\sum ((F_i - O_i) - (\bar{F}_i - \bar{O}_i))^2 \right]^{1/2}$$

Where an overbar represents the long term mean. Figure 3.2 illustrates the use of σ in model verification.

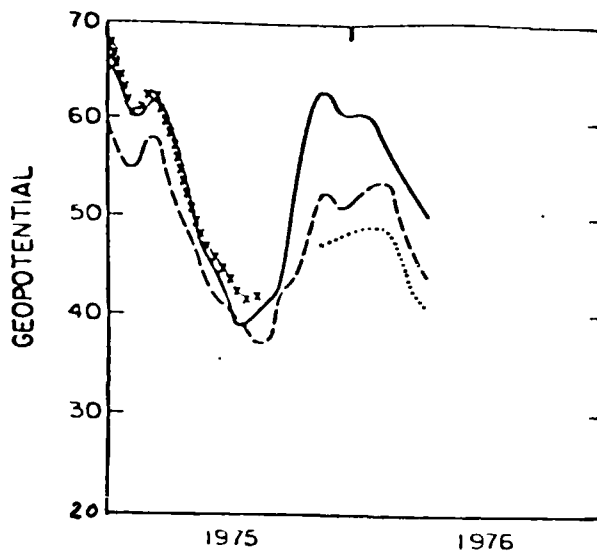


Figure 3.1 MSE error of the 36-Hour 500 MB geopotential forecast

Solid line = Canadian Filtered
Baroclinic Model
Dashed line = Canadian PE Spectral
(Rhomb. 20 truncation)
Dotted line = Canadian PE Spectral
(Rhomb. 29 truncation)
XXX = Canadian gridpoint PE

(From Daley, 1976.)

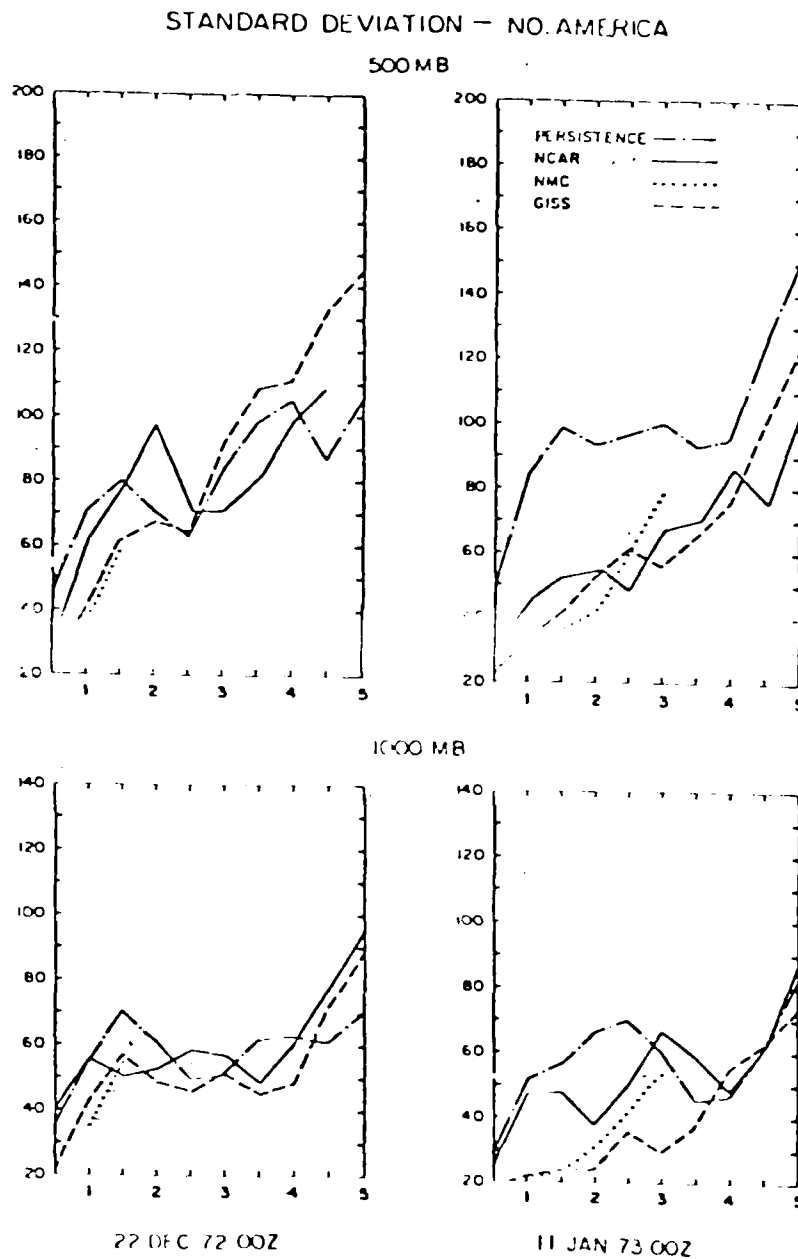


Figure 3.2

Standard Deviations of height for two 5 day model simulations. Forecast models are labeled.

(from Baumhefner and Downey, 1978)

The importance of σ in relation to the MSE is often overlooked in verification studies and systems. Many studies place much emphasis on the MSE values of hemispheric fields without any statement of the inherent variability within the system. Certain geographical areas are naturally more variable than other areas (both observationally and in model simulations). These naturally more variable areas influence the MSE value much more than less variable areas. For example if a model simulation has correctly forecast a cyclone in the Gulf of Alaska and incorrectly forecast two cyclones, one off of the coast of Japan and another off the east coast of the United States, the MSE value might yield a value indicative of an accurate forecast. This might occur because the Gulf of Alaska region is naturally more variable than the other areas. Therefore the variance of the individual grid points must be known before any significant degree of emphasis is placed upon the MSE measure. This requires a data base to be built up, enabling the variance to be calculated. Once the variance is calculated, the problem could be avoided by normalizing the data (e.g. removing the mean and dividing by the standard deviation). Another way to avoid the problem is to regionalize the calculation of the MSE value. Regionalizing the measure reduces the dependence upon developing a large data set yet still insures that the variance is relatively uniform throughout the domain of the MSE calculation.

The MSE measure is commonly applied to the wind vector as follows;

$$\text{Vector Wind error} = [(U_F - U_O)^2 + (V_F - V_O)^2] = \text{VWE}$$

and

$$\text{MSVWE} = \frac{\sum_i \frac{(\text{VWE})^2}{N_i^2}}{A},$$

where U and V are the wind components.

The MSE and σ values, as defined here, are very easy and inexpensive to calculate. Usually these measures can supply a good set of statistics usable in verifying model forecasts. However one should not over emphasize the importance or interpretation of these measures unless the natural variability is examined as discussed above and/or the score is regionalized.

Another disadvantage of these measures is that they tend to conceal errors in motion and intensity of synoptic scale features.

Through the nature of the calculation of these measures, the final value can be sensitive to any data smoothing or filtering. It is necessary to know of any pre and post processing as well as the numerical damping within the model.

Another disadvantage of the score as defined here is that they are not spectral. However these scores can be calculated for data fields broken down into the spectral domain (Arpe et.al., 1976).

3.1.1.2 Correlation

The sample correlation coefficient is a generalized measure of a relationship between pairs of variables from two samples. We define the sample correlation coefficient between pairs of forecast values, F_i , and observed values, O_i as:

$$r = \frac{N \sum F_i O_i - (\sum F_i)(\sum O_i)}{(\sum F_i^2 - (\sum F_i)^2)^{1/2} (\sum O_i^2 - (\sum O_i)^2)^{1/2}}$$

The sample correlation value is easy to compute and inexpensive in terms of operations within a computer. However it has a number of disadvantages. The r value is often subject to interpretation errors due to attaching too much significance to the correlation calculated from a small sample.

The sample correlation coefficient is also influenced by trends in the data. It is therefore recommended that the

calculation of r be done using departures from the normal (i.e. climatology) (Brier and Allen, 1951).

It is also very difficult to justify any statement concerning whether an actual relationship exists based on the correlation value. Freund (1972) describes the random variable, z , defined as

$$z = \frac{\sqrt{n-3}}{2} * \ln \frac{(1+r)(1-\rho')}{(1-r)(1+\rho')}$$

which has the standard normal distribution. This value can be used to test the hypothesis that the actual correlation is equal to ρ' versus the hypothesis that ρ does not equal ρ' . For example the value, z , will be used to evaluate whether a correlation value of .35, calculated for a particular sample, is significantly different from 0. In this case, ρ' would be 0 in the equation above. The difficulty in making objective statements concerning the correlation value is the major disadvantage to using this measure in model verification.

3.1.1.3 Brier Score

A final general measure of forecast accuracy is the Brier score (Brier, 1950). If we have n occasions of an event which can occur in any one of r possible classes, we define the probability of the occurrence during the i th occasion and the j th class as f_{ij} such that

$$\sum_{j=1}^r f_{ij} = 1, \quad i = 1, 2, \dots, n$$

The Brier score is defined as

$$P = \frac{1}{n} \sum_{j=1}^r \sum_{i=1}^n (f_{ij} - E_{ij})^2$$

where E_{ij} is 1 if the event occurred in class j and 0 if not.

This score can be applied to a number of forecast events. However this value is intended for use in measuring probability forecasts. Therefore this measure is not widely used for model verification. It is a useful measure for verifying probability forecasts which are based upon the model prognosis. This is an indirect verification of the model accuracy.

3.1.2 Graphical Methods

We define graphical methods of model verification as the representation of actual model parameters or statistical measures in various graphical configurations designed for highlighting certain features of the data.

3.1.2.1 Difference Fields

The graphical display of differences between a forecast and observed fields is a useful way of viewing errors in a model simulation (Baumhefner and Downey, 1978). An example of a difference field is shown in figure 3.3.

It is obvious that the difference field is easily computed and is helpful in displaying discrepancies in the intensity and position of synoptic scale features. A series of difference maps spanning various forecast intervals is helpful in determining errors in the movement of synoptic scale features.

The difference map provides a quick look at a model's performance. However it's necessary to point out a number of features before too much emphasis is place upon the difference field. As discussed in reference to the MSE it is necessary to have an indication of the variance of the individual points. This is necessary to evaluate if the difference between forecast and observation is actually different than the natural variation in the fields. A two tailed "t" test can be used to assess the significance of the difference fields. This is defined as:

$$t = \frac{F_i - O_i}{s \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$$

$$\text{where } s^2 = \frac{(n_1 - 1) s_1^2 + (n_2 - 1) s_2^2}{n_1 + n_2 - 2}$$

and n_1, n_2 are the number of data fields per group and s_1^2, s_2^2 are the sample variances for each group respectively. However it is necessary to have a data base or climatology of model forecasts and verifying analyses established. It is also necessary to combine cases which represent similar synoptic situations. For example the difference between forecast and observed Cape Hatteras Lows could be examined by combining a number of cases over a given period of time. This will allow for specific objective statements to be made concerning which areas are significantly different between forecast and verifying analysis.

3.1.2.2 Longitude-time plots (Hovmöller diagrams)

The Hovmöller diagram is a useful graphical technique for viewing the time evolution of the forecast fields. An illustration of this is shown in figure 3.4. Usually a specific latitude band is chosen to represent the time development of long and shorter scale features.

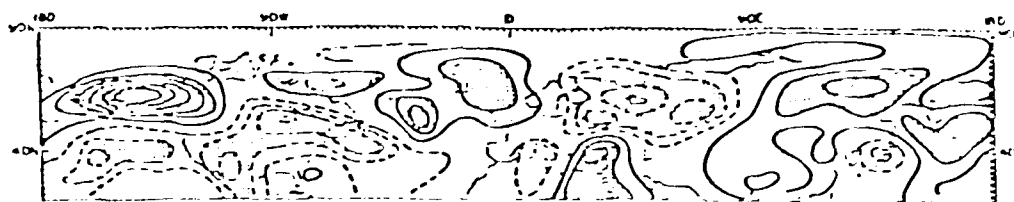
The Hovmöller diagram is particularly useful for displaying errors in the phase speed of synoptic scale features.

A further application of the Hovmöller diagram is to use the diagram in connection with spectrally decomposed data (figure 3.5). This is very useful for distin-

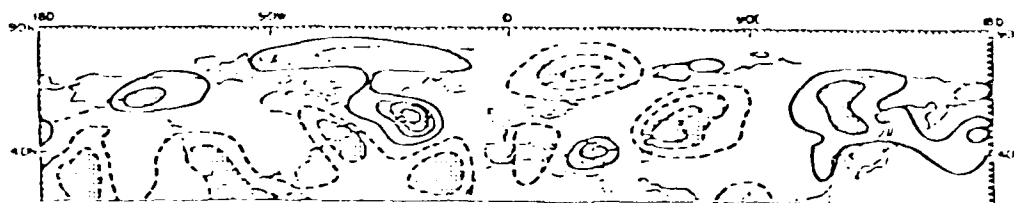
500 mb GEOPOTENTIAL
(FORECAST - OBSERVED)

11 JAN 73 00Z

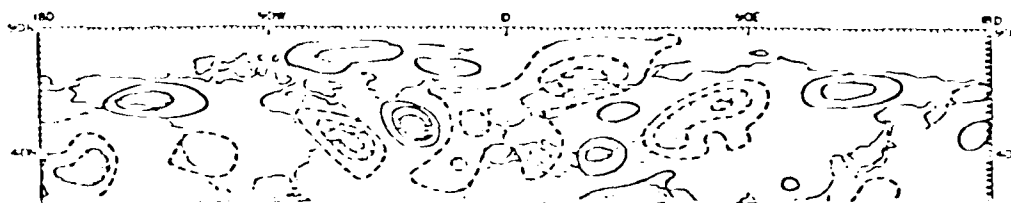
72 HR FORECAST



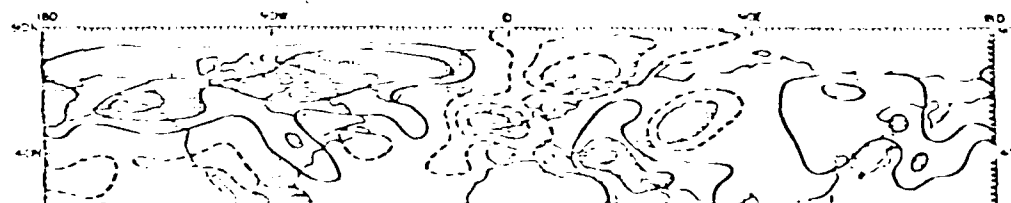
OBSERVED



NCAR



GISS



NMC

Figure 3.3

Difference fields of geopotential at 500MB for the observed change and three forecast (labeled) error fields for the 72 hour period beginning at 11 Jan.73 00Z. Solid lines indicate a positive value, negative values are denoted by a dashed line (from Saurhefner), 1976).

guishing the time development of planetary and shorter scale waves. However this can be lengthy and costly in terms of computing time. For this reason the Hovmöller diagram may not be considered a basic technique when used in connection with spectrally analyzed data. However in many cases the data is automatically analyzed spectrally at large forecast centers.

3.1.2.3 Zonal averages and Meridional cross sections

Many types of model parameters and data can be expressed in terms of zonal averages and displayed graphically, representing the latitudinal distribution of the parameter. This is illustrated in figure 3.6. The latitudinal distribution can also be displayed using a meridional cross section (figure 3.7). These graphical displays are quite helpful in viewing a latitudinal distribution of forecast errors. For instance an error in position and intensity of the jet stream seen in figure 3.7.

A similar graphical display is a time/height graph of model parameters or statistical measures. The time/height evolution of the RMSE for temperature over a two week simulation conducted by Miyakoda et.al (1972) using a hemispheric model at the Geophysical Fluid Dynamics Laboratory (GFDL) is shown in figure 3.8.

These types of graphical displays are quite useful for evaluating a model's performance. These figures are also easily constructed. In many cases the actual model results can be used with no further analysis necessary.

3.1.3 Other Specific Measures

There are a number of measures, some of which are statistical, which have been applied to specific meteorological parameters or forecasts. Measures described in this section are primarily used for precipitation verification, pressure verification or other forecasts based upon model output (e.g. cloudiness).

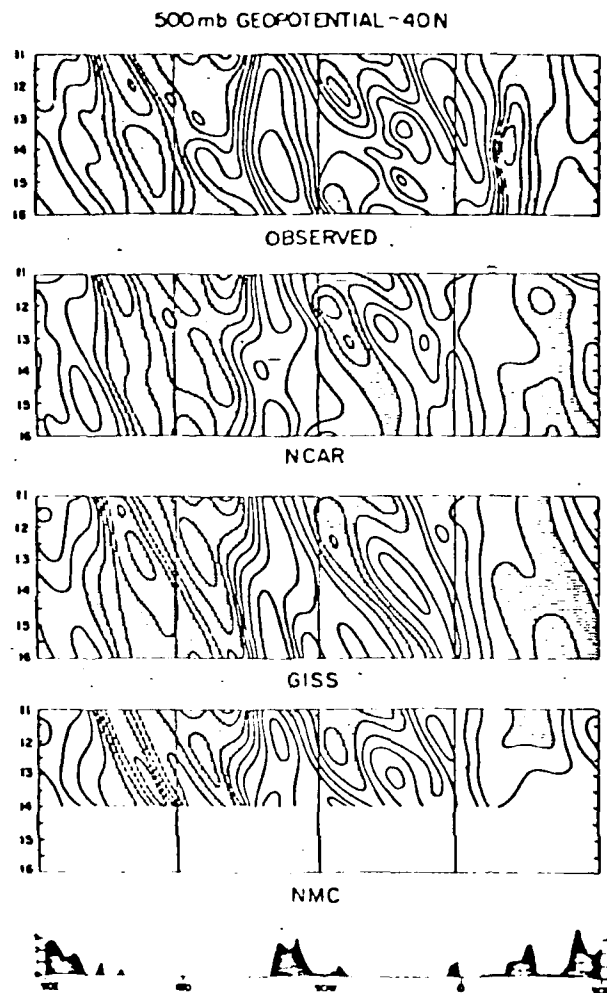


Figure 3.4

Hovmöller diagram of forecast and observed 500 MB geopotential at 40 N for 11 Jan. 1973. Time (in days) is the left ordinate. Longitude is on the bottom. Forecast models are labeled. Shaded values are below 5460 m with a contour interval of 60 m. (from Baumhefner and Downey, 1978)

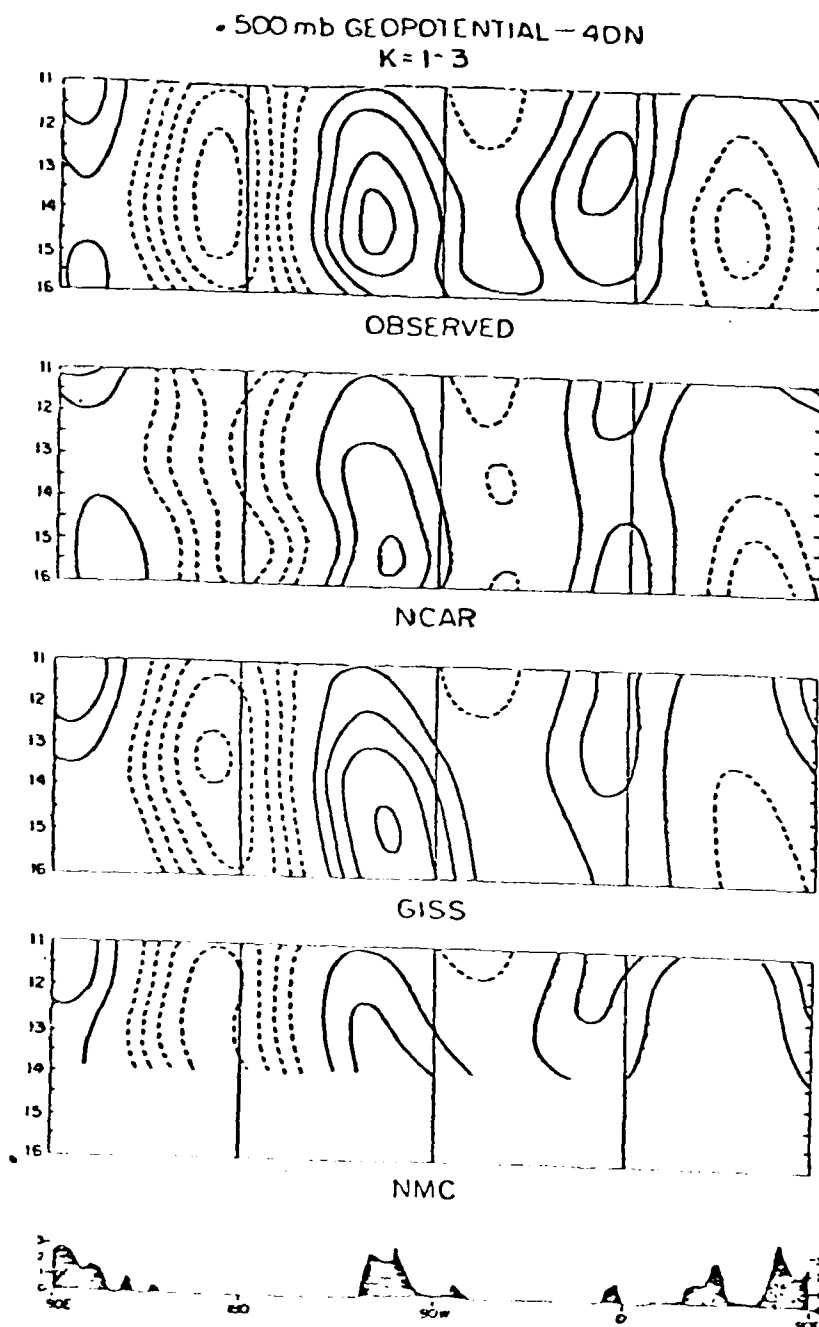


Figure 3.5

Hovmöller diagrams for wave numbers 1-3 of the 500 MB field at latitude 40°N beginning at 00GMT 11 Jan 1973 for the NMC analysis and three labeled models (from Baumhefner and Downey, 1976)

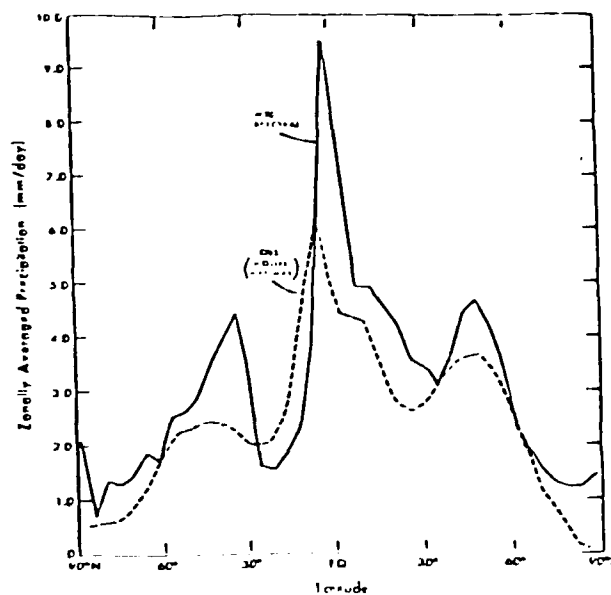


Figure 3.6

Latitudinal distribution of precipitation, forecast and observed (dashed line). The profiles are averaged over the period March 6 - March 15, 1973. Model forecast made with the GFDL spectral model. (Gordon, 1976)

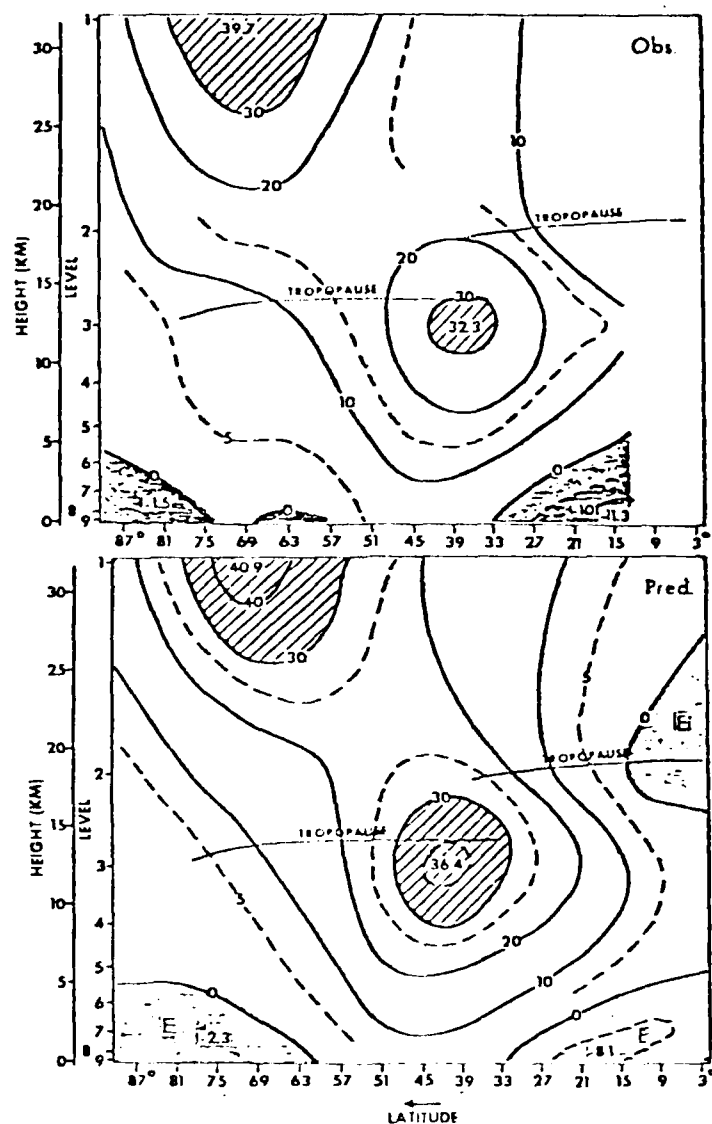


Figure 3.7

Meridional cross section of observed and predicted zonal winds (m/s). Forecast was made with the GFDL PE model and data is averaged over 10 days of each simulation of 12 different cases, all in January.
(from Miyakoda et.al., 1972)

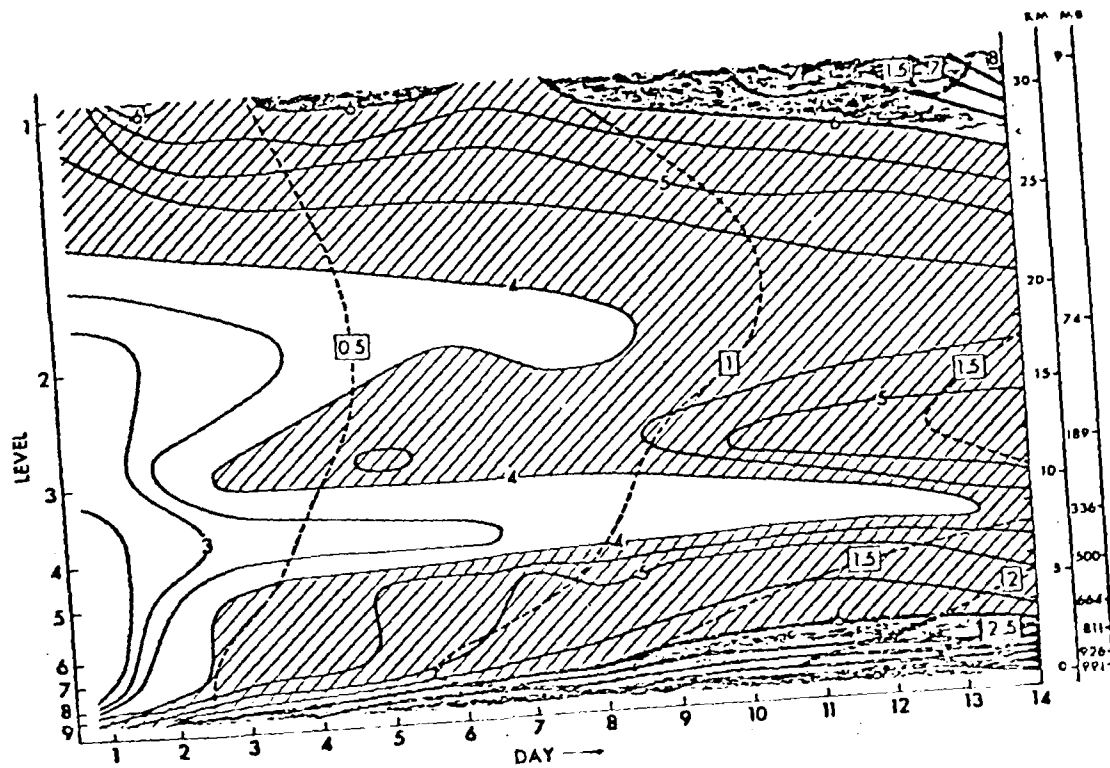


Figure 3.8

Time/height evolution of the RMSE temperature error over a two week simulation.
(from Miyakoda et.al., 1972)

3.1.3.1 S1 Score

The S1 score is used as a measure of skill in pressure forecasts. The score is defined as

$$S1 = 100 * \frac{\sum |E_G|}{\sum |O_G|}$$

Where E_G is the error in the forecast pressure difference between selected points at different locations and O_G is the observed or forecast pressure differences, whichever is larger.

An illustration of the S1 score is presented in figure 3.9. This score is easily computed but has the same disadvantage as discussed in connection with the MSE in that it is sensitive to data smoothing and hides errors in motion and intensity.

3.1.3.2 Threat Score

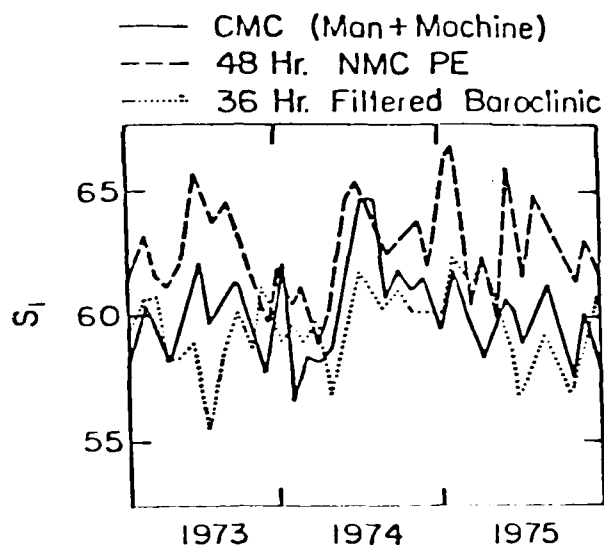
The threat score is used to measure the relative frequency of correctly forecasting an event in an area in which the event was a threat. A more formal definition is,

$$\text{Threat Score} = \frac{\text{Area Correct}}{\text{Area forecast} + \text{area observed} + \text{area correct}}$$

This score is illustrated in figure 3.10.

The threat score is used mainly in verifying precipitation forecasts and is applied to a specific region where the probability of an event occurring is large. A possible application of this measure is for use in verifying forecast events when the forecast is made by a meteorologist who uses the numerical prognosis for a guide. This would be an indirect verification of the model forecast.

This is an easy measure to compute,



36-Hour MSL S_1
Canadian Region

Figure 3.9

S_1 score comparison between the Canadian Meteorological Center Models and the NMC PE. The score was computed over the Central Canadian Region.

(From Daley, 1976.)

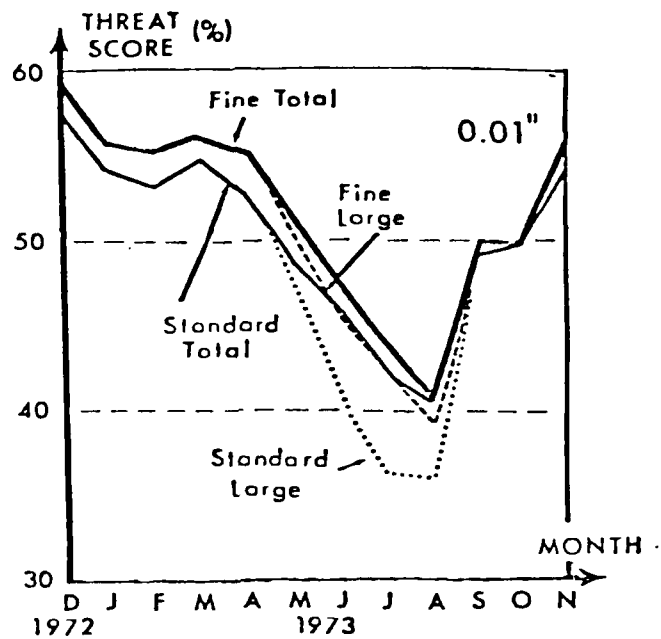


Figure 3.10

Threat Score calculates averaged for all North American stations for forecasts made with the Canadian Meteorological Center's standard fine mesh baroclinic models. (From Daley, 1976)

3.1.3.3 Prefigurance, Post Agreement and Heidke Skill Score
 Prefigurance and Post Agreement are defined in terms of contingency tables:

		Forecast Class		
		F ₁	F ₂	
Observed Class	C ₁	O ₁₁	O ₁₂	R ₁ = O ₁₁ + O ₁₂
	C ₂	O ₂₁	O ₂₂	R ₂ = O ₂₁ + O ₂₂
		L ₁	L ₂	N = TOTAL

Prefigurance is defined as the extent to which forecasts give an advanced warning of an event, given as

$$O_{ij}/R_i.$$

Post agreement is defined as the "percent right" given as

$$O_{ij}/L_i$$

These measures are also commonly applied to preceipitation events.

The Heidke skill score is easily computed from the contingency table as

$$S = \frac{F - E}{N - E}$$

Where N is the total in all cells, F represents the sum of cases in the correct forecast cells and E is the sum of cases in the incorrect forecast cells. This measure will vary between 0 and 1, with 1 being a perfect forecast.

As discussed in relation to the threat score, these measures are easily computed and interpreted but their utility is limited.

3.1.4 Summary of Basic Techniques

As discussed above, there are a number of basic techniques applicable to model verification. We can summarize by stating that all of the techniques discussed are easily computed, inexpensive to compute and can quickly provide a good comparison of a model forecast and verifying analysis. However many times too much emphasis is placed on the significance of results without insuring that the methods or interpretations have a sound statistical basis. In order to insure this, it is necessary to have a data base established, enabling a large number of samples to be used.

Therefore there is a trade off between the simplicity of the techniques and the validity of the results. These basic techniques can provide a very thorough and comprehensive evaluation when applied correctly.

3.2 Detailed Techniques

Detailed techniques of model verification are not always statistical but usually provide more insight into the physics of certain systematic errors occurring within numerical forecasts. We can discuss these detailed techniques in terms of analysis techniques (e.g. spectral analysis) which re-analyze a basic data field enabling a detailed examination of the model when even the basic techniques described above are used. We can also discuss other more detailed techniques or studies (e.g. pattern recognition) which can be applied to a basic data set or re-analyzed data set.

3.2.1 Spectral Analysis

Spectral decomposition separates the components of the space time variance of a function of latitude and time

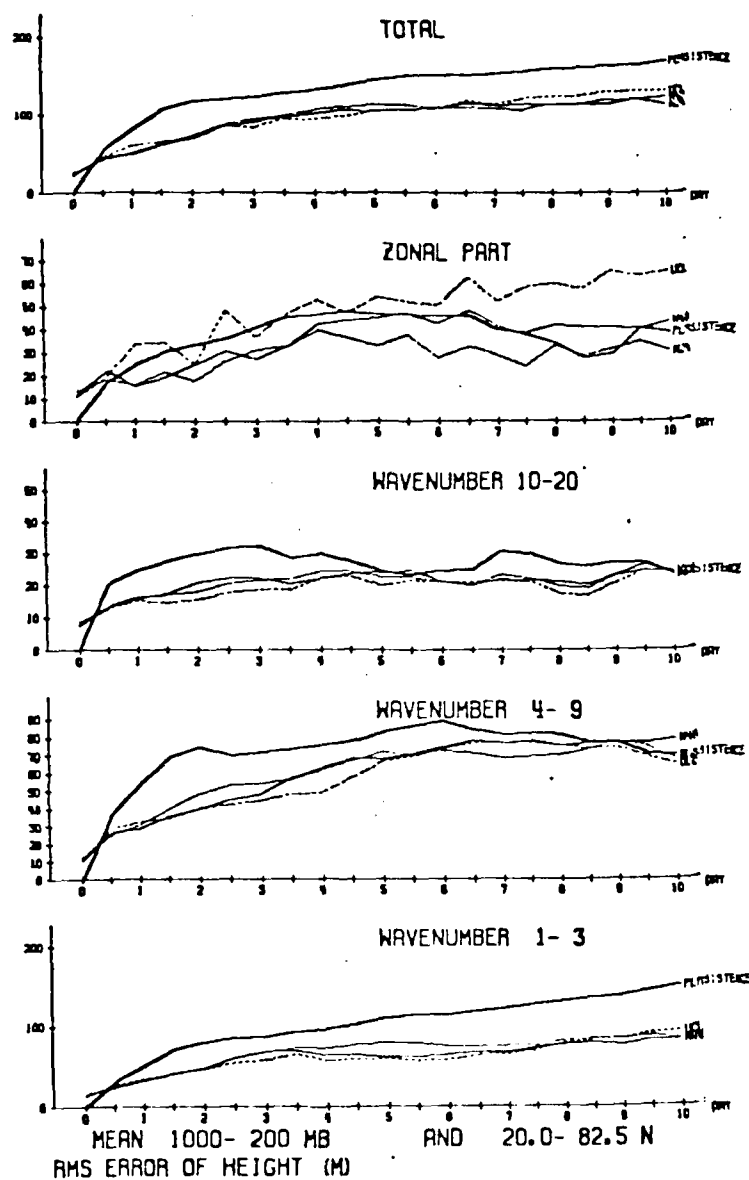


Figure 3.11

Spectral RMSE of height for a 10 day simulation using the UCLA GCM and the GFDL N24 and N48 models. N24 refers to 24 latitudinal grid intervals between the equator and pole while N48 represents 48 intervals. Simulation from 5/3/65.

(from Arpe et.al. 1976)

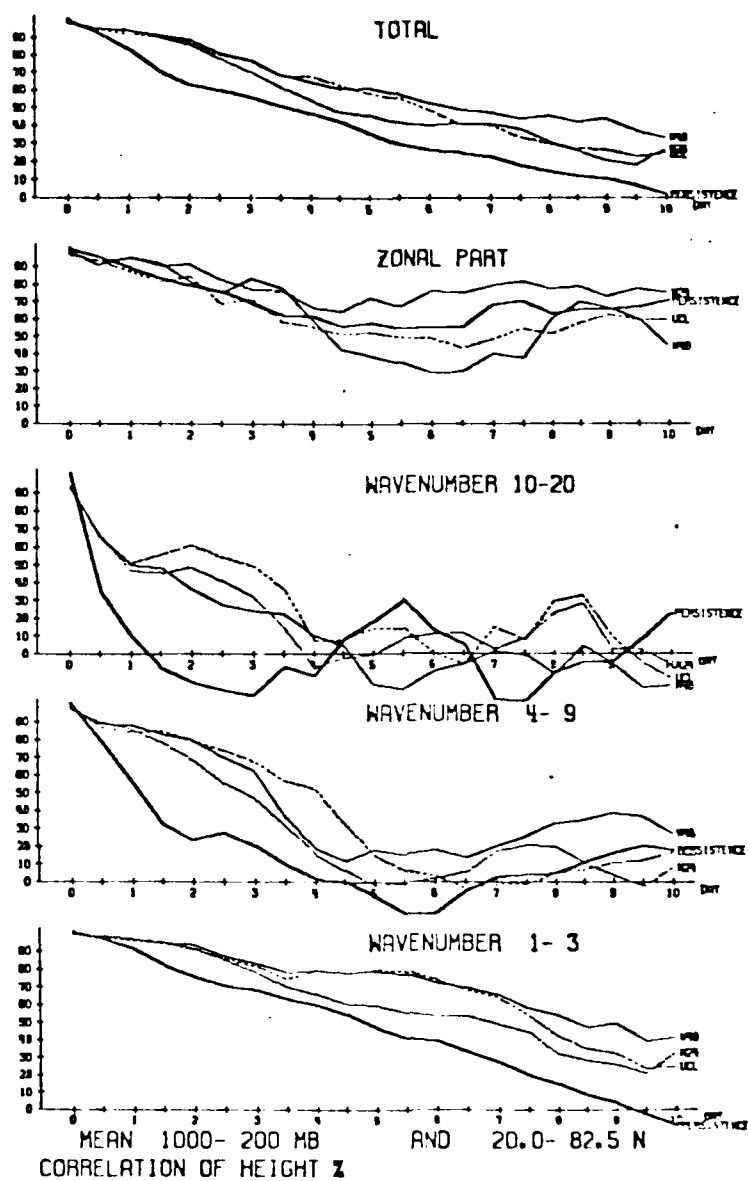


Figure 3.12

Spectral correlation of height for a 10 day simulation. Same models as in figure 3.11.

(from Arpe et.al. 1976)

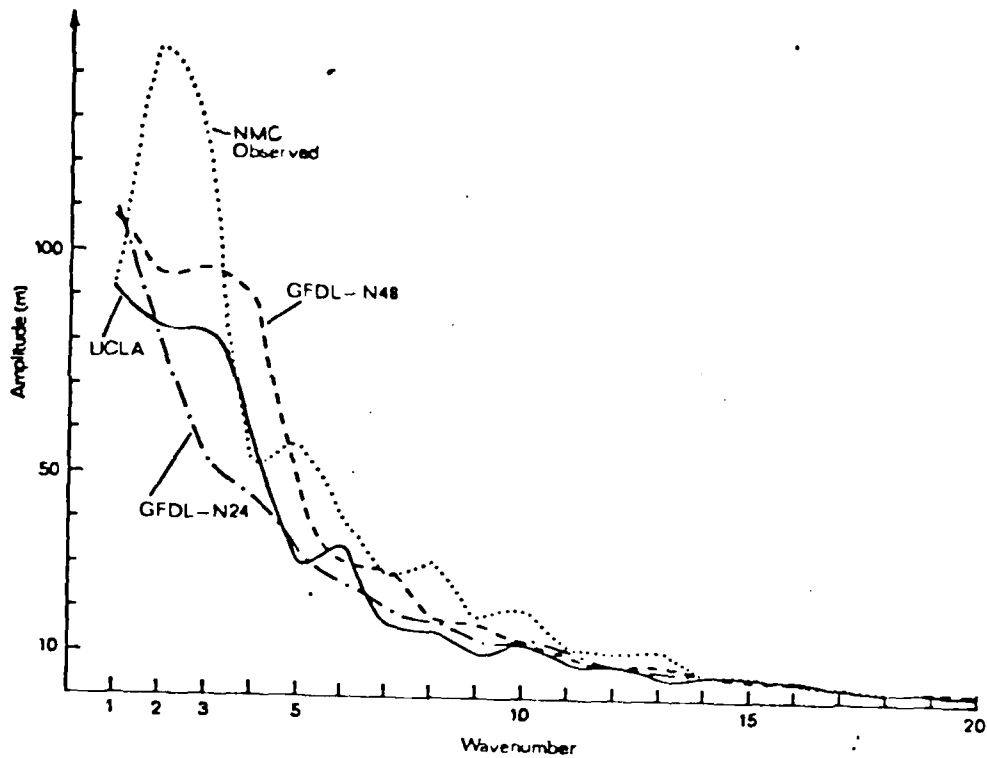


Figure 3.13

Amplitudes of geopotential height waves between 40 N and 60 N at 500 MB. Same models and simulation as in figure 3.11. (from Arpe et.al. 1976)

into a full wavenumber frequency expansion (Lorenz, 1967). Spectral analysis can be applied to a data field, thereby allowing techniques described under the basic analysis to be evaluated using spectrally transformed data.

After the data is analyzed in this manner most of the methods described as basic techniques can be used to obtain a set of statistics which discriminate in the wavenumber domain. Figure 3.11 - 3.13 illustrate the use of RMSE, correlation and a graphical display of wave amplitudes presented by Arpe et.al. (1976) in their analysis of several numerical models. These analyses enable the basic techniques to reveal the skill associated with the quasi stationary longwaves ($K = 1-3$), rapidly moving baroclinic waves ($K=4-9$) and shorter waves ($K=10-20$). Figure 3.6 illustrated the utility of the Hovmöller diagram when applied to spectrally analyzed data. There are other, more detailed, techniques which can be applied to spectrally analyzed data and will be described later. Even though the basic techniques such as RMSE can be applied to spectrally analyzed data the formulation of these parameters becomes much more complex (Arpe et. al 1976).

Spectral analysis can also be applied in the more traditional sense with respect to time series analysis. This would apply to the analysis of a forecast and analysis time series, say, of temperature at a particular location. Spectral analysis could then be used to locate differences in period and/or amplitude of the two series thus indicating forecast errors.

3.2.2 Quasi Lagrangian analysis

The application of quasi Lagrangian diagnostics to model verification has become more common (Downey and Johnson, 1978, Wash and Johnson, 1977). This method utilizes a conical shell around a cyclone or anticyclone which moves with the system. This allows computations and studies of budgets of mass, energy and momentum. The exact formulation is quite complex and the reader is referred to Johnson and

Downey (1975) for a detailed description of the technique. However the method is quite useful for representing data in a form where the interactions between a synoptic scale system and the larger scale environment can be analyzed for model simulations and compared to the actual situation.

3.2.3 Quasi Geostrophic Analysis

A further type of analysis available for detailed descriptions of model accuracy is the use of the concepts of quasi geostrophic theory (Holton, 1972). This analysis is applied to various models by Houghton and Irvine (1976). Representation of data in terms of the omega equation and tendency equations (Holton, 1972) allow investigations into the processes which govern the evolution and development of synoptic scale features. Application of the omega equation allows the evaluation of such parameters as; vertical motion, increase with height of positive or negative vorticity advection and thermal advection. The tendency equation evaluates the geopotential tendency, vorticity advection and the decrease with height of thermal advection.

We have now described three types of analysis which can be used to represent basic model and verifying analyses data fields in such a way as to allow a detailed examination of the physical and dynamical differences between the model and actual atmosphere. Spectral and quasi Lagrangian analyzed data can be used to examine budgets of specific quantities such as energy parameters. Quasi geostrophic analysis is a more self contained process in that this method provides specific quantities (defined by the omega and tendency equations) while spectral analysis, for instance, can be applied to many types of data fields as a tool for other techniques.

3.2.4 Regression Techniques

Similar to the RMSE and correlation coefficient regression analysis does not always lend itself to easy or

straight forward interpretation. In a regression type verification there are two basic cases (figure 3.14). Case A is the case of a perfect relation between model prediction and observation (O). This case is the line $P=O$. Case B is the imperfect relationship between P and O shown by the equation

$$O = A+BP.$$

In terms of case B two questions should be asked (Brier, 1975);

- i) Can this line be used for calibration?
- ii) Is this result useful in other areas?

In order to evaluate the regression scheme one must evaluate the ways errors can creep into a model. Brier (1976) has shown that the regression coefficient, B, is drawn away from the perfect case by the variance of error components associated with the model, M, and initialization, I, when the predicted value has the following formulation

$$\text{Predicted value} = \text{true prediction} + M + I.$$

It is necessary to identify the terms M and I in order to successfully evaluate and interpret the regression results. This is the main disadvantage to the use of regression in model verification.

3.2.5 Pattern Recognition

The technique of pattern recognition as applicable to model verification has been described by Somerville (1977) and is summarized here. This technique involves three steps:

- i) Representation of input data;
- ii) Extracting features;
- iii) Assigning patterns.

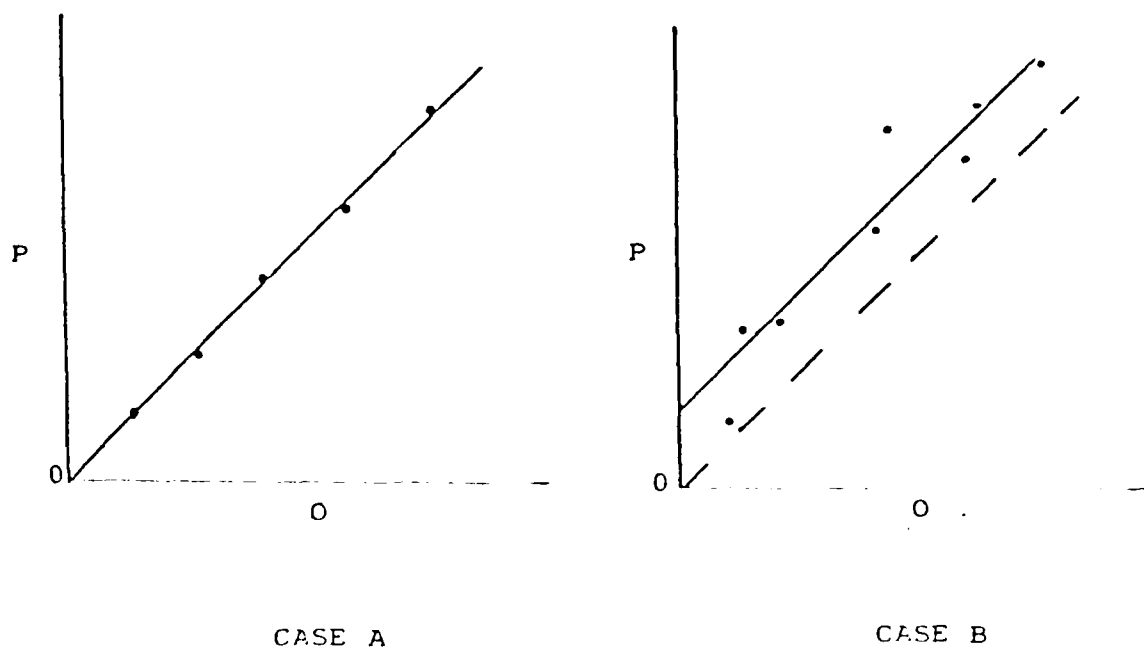


Figure 3.14 Two cases of a regression verification scheme:

Case A - Perfect model

Case B - Model is not perfect

Input data of forecast, F, and analysis, A fields can be arranged in matrix form where rows, j, represent the spatial variation of the data field and the columns, n, represent the temporal variation of the forecasts and analysis.

$$\underline{F} = \begin{bmatrix} F_{01} & \dots & F_{0j} \\ F_{11} & & F_{1j} \\ F_{n1} & & F_{nj} \end{bmatrix} \begin{array}{l} \text{- initial conditions} \\ \text{- forecast at time 1} \\ \text{- forecast at time n} \end{array}$$

grid points 1 3969 (63x63)

$$\underline{A} = \begin{bmatrix} A_{01} & \dots & A_{0j} \\ A_{11} & & A_{1j} \\ A_{n1} & & A_{nj} \end{bmatrix} \begin{array}{l} \text{- initial conditions} \\ \text{- verifying analysis at} \\ \text{time n} \end{array}$$

A difference field is defined as

$$\underline{D} = \underline{F} - \underline{A} = \begin{bmatrix} 0 & \dots & 0 \\ D_{11} & & D_{1j} \\ D_{n1} & & D_{nj} \end{bmatrix} \begin{array}{l} \text{- initial fields are} \\ \text{identical} \end{array}$$

Where $D_{0j} = 0$ represents a perfect initialization

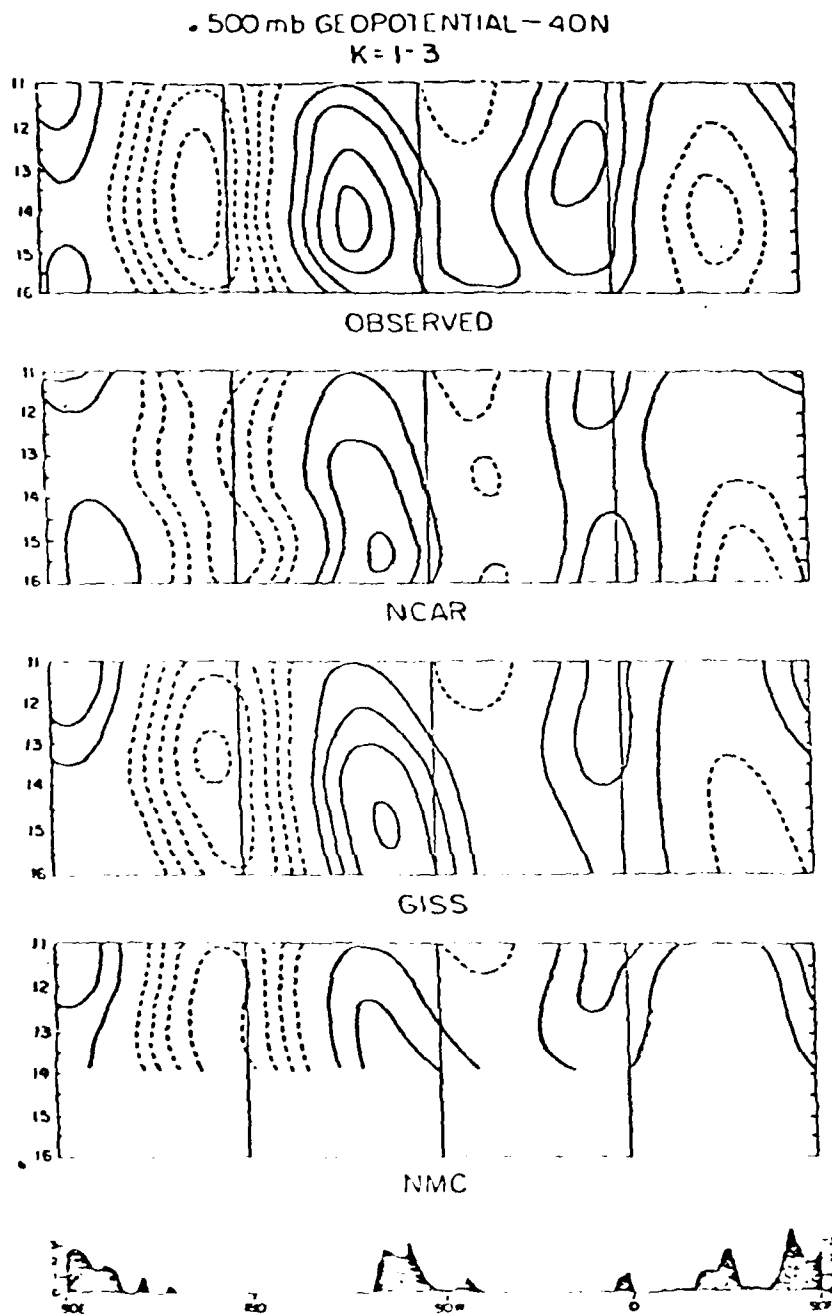


Figure 3.15 Hovmöller diagram used for Pattern Recognition as defined by Somerville (1977).
(from Somerville, 1977)

Somerville describes the feature extraction phase in terms of a Hovmöller diagram shown in figure 3.15. Defining the feature as the trough and ridge lines, figure 3.16 represents a schematic diagram of a feature matrix, M , for the analysis field. Once the feature matrix is obtained specific properties can be extracted. In this case Somerville extracts ridge and trough lines not present during initialization (figure 3.17). This pattern matrix for the analysis field, A_s , can be compared to other pattern matrices derived from the various forecast fields, F_s . This type of comparison can show the following:

- i) Features in A_s only, represent those not forecast;
- ii) Features in F_s only, represent those forecast but not observed;
- iii) Features in both A_s and F_s provide information on the time of entry into the forecast, amplitude of the wave and speed of the wave.

This technique provides a very comprehensive look at particular features of model forecasts and verifying analyses. The application discussed by Somerville and summarized here seems to have a great deal of utility in providing information on phase and amplitude errors of long waves and how these errors depend on time.

A variation of the pattern recognition techniques discussed here has been described by Holl and Cuming (1979) with respect to the FNOC model and analyses. Measures of Synoptic Similarity (MOSS) differs from Somerville's concept in that an objective measure is produced which represents the degree of similarity between data fields. Data fields are broken up into the following ranges of scale:

- i) SD - describing propagating cyclones and anti-cyclones;

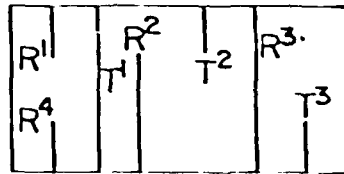


Figure 3.16 Schematic representation of the sparse matrix used to identify the ridges (R) and troughs (T) in figure 3.15. (from Somerville, 1977)

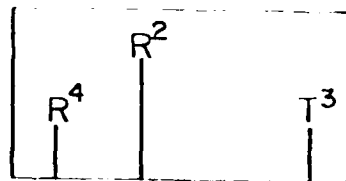


Figure 3.17 Schematic representation of the string matrix consisting of features not present in the initial conditions (from Somerville, 1977)

- ii) SL - describing the various centers of action (e.g. semipermanent highs and lows, long wave patterns);
- iii) SV - describing the planetary vortex scale.

Objective measures are obtained for the three ranges of scale and the 1000MB, 500MB and 1000MB - 500MB thickness fields. The measures are based on bit coding values of gradients and grid point values of the respective fields. The final value is obtained by comparing the number of matching bits, coded for the respective fields being compared.

This method is useful for providing an objective measure of similarity between forecast and verifying analysis. However a large number of samples would be needed to enable an objective interpretation of the resulting measures. This would be accomplished as a data base is built.

The two methods of pattern recognition described here each have distinct advantages for specific applications. The Somerville method seems applicable to spectrally analyzed data and less complicated than the method described by Holl and Cuming. However the latter method provides an objective measure while the former is more subjective (although parameters such as RMSE and σ can be incorporated into the process).

3.2.6 Diagnostic Studies

Diagnostic Studies are quite useful for comparing various higher order parameters of the model and real atmosphere. This technique is commonly applied to the energetics of the atmosphere. The main advantage is that this type of analysis will often reveal more subtle differences in the height and wind field than a more simple analysis. The main disadvantage to the analysis is the vertical velocity is often required to evaluate parameters such as the generation or dissipation of energy. Values of vertical velocity are usually readily available in model simulations, however, the ability to attain this parameter in the actual atmosphere is poorly defined. This prevents

degree of significance from being attached to values computed from the vertical velocity (Pearce, 1974).

The following energy parameters are commonly used in diagnostic studies applicable to model verification:

- i) Zonal available potential energy, AZ;
- ii) Eddy available potential energy, AE;
- iii) Zonal kinetic energy, KZ;
- iv) Eddy kinetic energy, KE;
- v) Conversion of AZ to AE, C(AZ,AE);
- vi) Conversion of KE to KZ, C(KE,KZ).

These values can also be calculated using spectrally analyzed data (Arpe, et. al., 1976). Formulation of these parameters are straight forward and are described by Lorenz (1967).

Budget equations can be found as follows (for the eddy terms);

$$\frac{\partial KE}{\partial t} = - C(KE, KZ) + C(AE, KE) - D(KE)$$

$$\frac{\partial AE}{\partial t} = C(AZ, AE) - C(AE, KE) + G(AE).$$

However these values are dependent upon the vertical velocity.

Figure 3.18 shows observed and forecast kinetic energy for a verification study conducted by Ward et.al., (1977). This figure illustrates the usefulness of this type of analysis in identifying the time variation and error in such parameters as the jet stream.

Diagnostic studies are also used in connection with quasi Lagrangian analysis as mentioned above. Wash and Johnson (1977, 1979) used quasi Lagrangian diagnostics to examine the budgets of mass and angular momentum in model and actual cyclonic systems. Figure 3.19 illustrates this type of study for model verification.

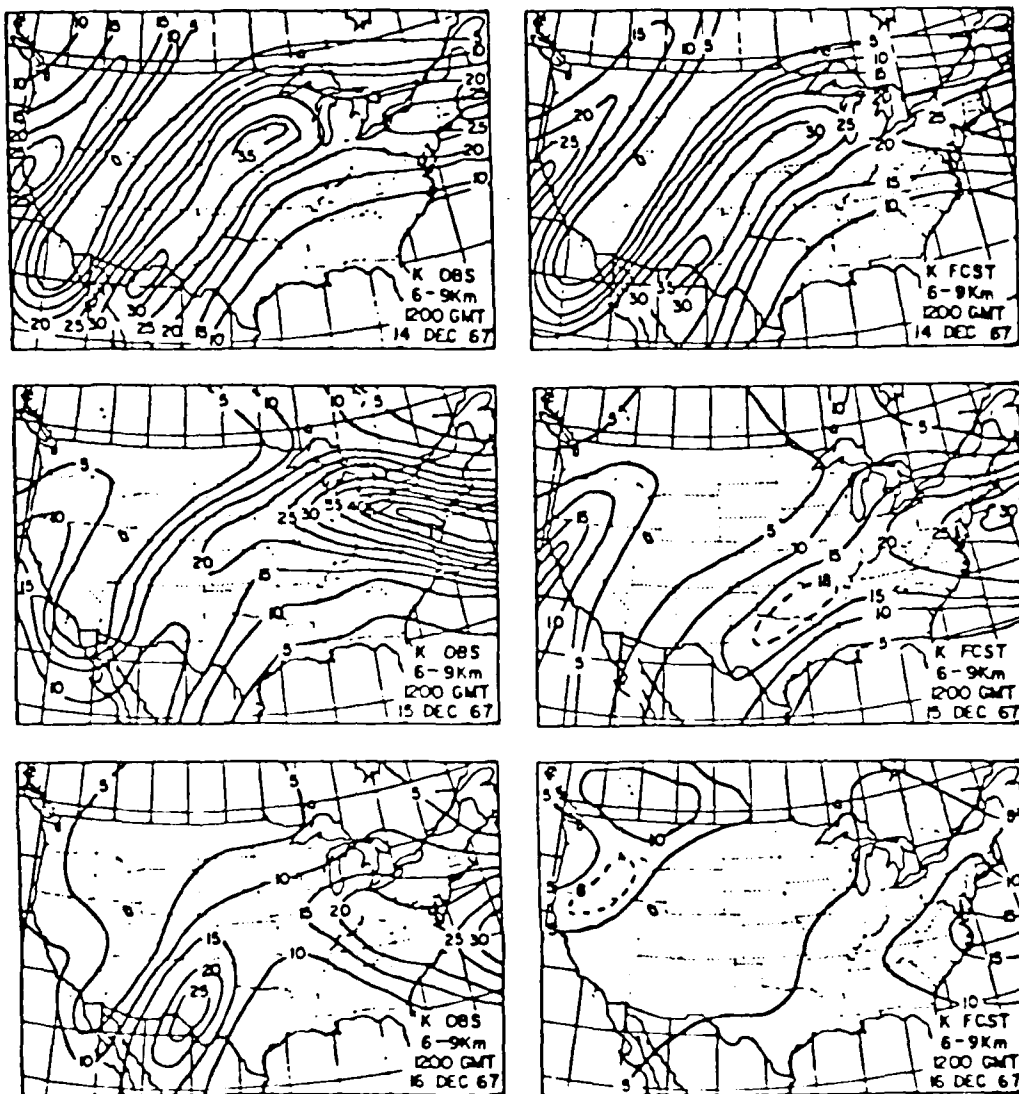


Figure 3.18

6.9 km observed and forecast kinetic energy
for 1200 GMT, 14-16 December, 1967.
Simulations conducted with the 6 layer
NCAR GCM.

(from Ward et.al., 1977)

Diagnostic studies can be very helpful in highlighting differences between forecasts and verifying analyses. Budget studies of diagnostic quantities, requiring the use of vertical velocity are not as reliable as examining the spatial variation of a single parameter. A drawback of the quasi Lagrangian analysis is that the vertical velocity is required for calculations. Often a comprehensive diagnostic study of energy or other parameters can indicate errors within more simple or lower order model parameters which might not have been evident using a simple RMSE, σ , or difference field. The diagnostic study is probably the most complete detailed analysis that can be applied to a situation or region for model verification.

3.3 Specific Techniques

All of the techniques described under the basic and detailed classes have been applied many times to mid latitude synoptic scale features present in numerical model forecasts. However evaluations of a model's performance in the tropics is not as simple. For instance, use of an RMSE, σ or difference value for evaluating a model's performance in the tropical regions is almost totally useless. The small natural variability in parameters such as height and temperature constrain any values of RMSE or σ to no meaningful and statistically valid interpretation.

Verification of a model's forecast in the tropical regions is best accomplished by verifying certain events and/or large scale processes which are clearly defined within the region or which link the mid latitude and tropical circulations. Examples of such events and/or processes are the monsoon circulation, location and intensity of the subtropical jet stream and the position of the Inter Tropical Convergence Zone (ITCZ). The monsoon is an event where the nontropical circulation can dominate or force the more tropical circulation therefore allowing the traditional verification schemes to be used (Payne, 1979).

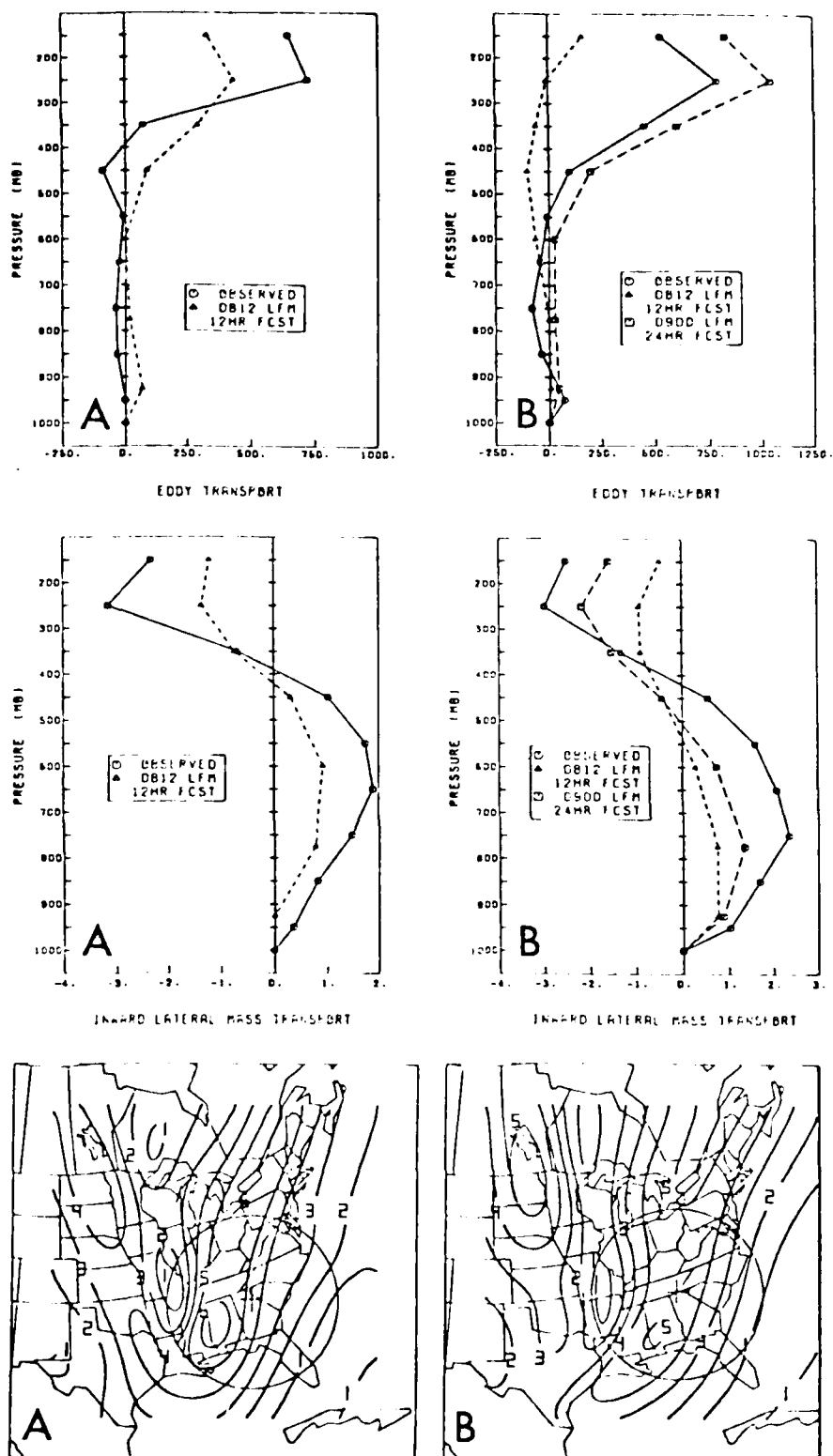


Figure 3.19

Figure 3.19

(Top) vertical profiles of eddy mode of angular momentum transport (10^{15} kg M^2 /sec.) Model predictions were made with the NMC LFM model. Observed and forecast times are marked. The cyclone is positioned over the midwestern United States.

A. 0000 GMT 9 Oct.

B. 1200 GMT 9 Oct.

(Middle) Profiles of azimuthally-averaged inward mass transport (10^{10} kg/sec). (lower) Comparison of isotach patterns (M/sec) for observed (A) and 12 hr prediction by 0812 LFM (B). (from Wash and Johnson, 1979)

The position of the subtropical jet can be identified readily in model fields where "data" is available at all positions for all times, however positioning of this jet can be difficult in the actual atmosphere where observations are sparse.

Verification of the ITCZ can be attained by use of satellite imagery however this is often difficult to resolve. Pattern recognition is also useful in identifying the ITCZ.

3.4 Summary

The most commonly used methods used in model verification have been outlined in this section. Methods were discussed under the context of the basic and detailed analysis described in Section 2.

Basic techniques are largely statistical in design and are primarily used to provide a quick general comparison between a model forecast and verifying analysis. The most important point to remember when using these techniques is that often important interpretations are applied to results obtained with these techniques without considering points such as sample size or the natural variability within the field which affect the statistical validity of the results.

Detailed techniques discussed here require more computing time and evaluation than the more basic techniques. However they provide more insight into the physics of discrepancies between model forecasts and verifying analyses. The most useful results can be obtained when the two types of techniques (basic and detailed) are applied in such a way as to compliment each other. This will provide a very comprehensive look at numerical model accuracy and quality.

3.4.1 Advantages, Disadvantages of techniques

We can summarize the various techniques and measures in terms of the advantages and disadvantages of their use.

3.4.1.1

Basic Techniques

i) MSE

Advantages - The MSE is a very simple and inexpensive to compute. It is also applicable to a large number of parameters and model types. This type of measure is currently used at FNOC for the evaluation of the existing operational forecast model.

Disadvantages - The MSE should be evaluated in connection with the σ of the data field. The nature of the MSE dictates that it is very sensitive to the variance within the domain of the data field. It is necessary that the data be normalized or the score regionalized before a great deal of emphasis is placed upon the interpretation of the MSE value.

The MSE is insensitive to errors in the motion and intensity of the synoptic scale features.

ii) Correlation

Advantages - Like the MSE value, the correlation coefficient is very easy and inexpensive to compute. It is also applicable to a large number of parameters. A current software package at FNOC is capable of evaluating the correlation value for input data fields.

Disadvantages - It is very difficult to make objective statements concerning the significance of the correlation measure. There is a method which can evaluate a hypothesis test concerning the significance of a correlation value however it is not commonly used. Large data samples are necessary in order to objectively evaluate the significance of a correlation value.

iii) Brier Score

Advantages - The Brier score is easily computed. This score is applicable to products derived from numerical forecasts.

Disadvantages - The Brier score is mainly intended for measuring the accuracy of probabilistic forecasts. It is not widely applicable to verification of numerically produced forecasts.

iv) Difference Fields

Advantages - The difference field is straight forward to compute and plot using existing software products at FNOC. A number of particular difference fields are already computed at FNOC. The difference field is useful in displaying discrepancies between intensity and position of synoptic scale features.

Disadvantages - It is often necessary to make an objective statement regarding which areas are significantly different between two fields. This is necessary to determine which areas are different in terms of the signal of the field rather than observing differences in the noise of the two fields. This requires evaluating the "t" statistic at the individual grid points giving a field of "t" contours which actually indicate the significant differences between two fields. This is a simple procedure in terms of computation but requires a data base in order to give a valid number of degrees of freedom for evaluating the confidence level of the "t" statistic.

v) Hovmöller diagrams

Advantages - The Hovmöller diagram is useful for indicating various phase differences in synoptic scale features between forecast and analysis. Depending upon the type of diagram, amplitudes may also be indicated in the graph. Diagrams indicating phase errors between trough and ridge lines are capable of being produced using existing software package at FNOC. This package is capable of displaying the graph in the spectral domain.

Disadvantages - Interpretation of a Hovmöller diagram is not always straightforward. Interpretation can be made easier by differencing diagrams representing forecast and verifying analysis respectively. Arranging a diagram which shows ridge and trough lines for analysis and forecast together will make interpretation easier.

vi) Meridional cross sections

Advantages - Meridional cross sections are capable of displaying discrepancies in large scale atmospheric features. Cross sections of the wind component will enable one to verify positions of the jet streams for instance.

Disadvantages - Meridional cross sections are mainly used for general circulation models which frequently are used for simulating large scale climate rather than producing operational synoptic scale forecasts.

Interpretation of individual cross sections may be hard to interpret, as mentioned in connection with the Hovmöller diagram. Combining information on respective forecast and verifying analyses cross sections in terms of difference will aid interpretation.

vii) SI Score

Advantages - The SI score is easily computed from grid point data. This measure is currently in use at FNOC for evaluation of the current operational forecast model. The SI score is very useful in verifying the winds of a forecast and analysis field.

Disadvantages - The SI score is very sensitive to any data field smoothing or filtering taking place within the model and analysis framework. A complete understanding of the smoothing is necessary before any great deal of significance is placed upon SI score results.

viii) Threat score

Advantages - The threat score can have a great deal of utility with respect to representing results of verification methods in a very simple manner. The threat score is easily computed and can also be used to verify events which are not explicitly forecast by the numerical model but whose forecast is dependent upon the numerical prognosis of other fields (e.g. visibility, ceiling)

The threat score has great utility in expressing results of detailed, say, budget studies of energy parameters when the results are arranged in an appropriate manner. This will provide a simple measure which may be

easily interpreted and represents a more complex analysis which is not always easily understood.

Disadvantages - It is not always possible to make an objective statement concerning the significance of a threat score value. The threat score value will be very dependent upon each individual case or study.

ix) Contingency table related measures

Advantages - As discussed above in reference to the Threat score contingency tables have great utility for representing more detailed analysis results in a straight forward and concise manner. Post agreement and prefigurance can attach a simple measure which is easily interpreted to an analysis which originally is quite complex and detailed.

Tabulation of these measures are straightforward and inexpensive in terms of computation time and use of resources.

Disadvantages - It is often difficult to assign an objective degree of confidence to values of post agreement and prefigurance. The chi-squared test can be applied to the contingency table, however it is necessary to have an adequate sample in order to insure an acceptable number of degrees of freedom necessary for a valid application of the chi-squared test.

3.4.1.2

Detailed Techniques

i) Spectral analysis

Advantages - Spectral analysis is extremely useful for analyzing the different wave modes of synoptic scale features. The capability

of spectrally analyzing data already exists in a current software package at FNOC. Spectrally analyzed data can be further investigated using many basic techniques.

Disadvantages - The analysis of spectral data, using other basic techniques (e.g. MSE, correlation) is subject to problems of interpretation as described with respect to the techniques mentioned above.

ii) Quasi Lagrangian analysis

Advantages - Quasi Lagrangian analysis is an extremely useful method for examining the interaction of synoptic scale features with the larger scale atmospheric environment. This type of analysis can provide a great deal of insight into the physical mechanisms of the model.

Disadvantages - Quasi Lagrangian analysis is a very complex procedure and would require a development period before it could be implemented at FNOC.

Interpretation of quasi Lagrangian analysis results could be quite involved. It is necessary to try and arrange results in a manner which would enable the assignment of simple measures (e.g. threat score) to aid in the interpretation of the analysis.

iiv) Quasi Geostrophic Analysis

Advantages - Quasi Geostrophic analysis is useful for examining parameters which are important for the physical development of synoptic scale features.

Disadvantages - Like quasi Lagrangian analysis, quasi geostrophic analysis is quite complex and would involve some development before it can be implemented at FNOC.

iv) Regression Techniques

Advantages - Regression analysis can statistically relate various physical parameters to synoptic scale features. It is useful for examining the difference in these relationships between forecast and analysis parameters respectively.

Disadvantages - It is often difficult to assign any statistical confidence to regression results. Regression is most often used in numerical weather forecasting as an adjustment to forecasts based upon statistically measured relationships.

v) Pattern Recognition

a) MOSS

Advantages - The MOSS technique of pattern recognition provides a measure which represents the degree of similarity between two data fields. This method discriminates between the different wave modes (e.g. long waves, short waves through a simple and straightforward method. This procedure is currently in use at FNOC for evaluation of the existing forecast model.

Disadvantages - It is difficult to attach any objective degree of confidence to the MOSS

derived scores. This leads to possible ambiguity in the interpretation of the MOSS score.

b) Pattern Recognition of Somerville (1977)

Advantages - This form of pattern recognition is useful for further analysis of spectral data. Results of this type of pattern recognition can easily be arranged in a schematic diagram which can be easily and quickly diagnosed.

Disadvantages - This form of pattern recognition is not currently in use at FNOC and would thus require a period of development.

vi) Diagnostic Studies

Advantages - Diagnostic studies are very useful for analyzing the physical parameters of synoptic scale features in the atmosphere. This type of analysis can supply a comprehensive view of a model's performance with respect to many types of parameters for many types of atmospheric systems.

Disadvantages - Diagnostic studies often require data which is not easily attained in the actual atmosphere. The results of these studies are not easily interpreted. The results can be arranged, however, in a format which could be re-analyzed in terms of a threat score or contingency table, thus enabling a quick and easy interpretation. Diagnostic studies are not routinely conducted at FNOC and would require a period of development

3.4.2

Conclusions

The most important feature to consider in evaluating the utility of various techniques is the ability to interpret the analysis results in a timely and accurate fashion. The more basic techniques can easily be interpreted directly from the actual analysis. Other, more detailed techniques often require a re-analysis of the output which would ready the results for a timely interpretation. For instance, a budget study of energy parameters could be output in a time/space plot, allowing an evaluation of a type of threat score which would tell the viewer certain characteristics of the relative accuracy of the forecast. This type of output preparation and the cost, in terms of resources used and time spent, will be the main feature which will govern the utility of certain verification methods.

4.0 SYNOPTIC SCALE VARIABLE USED FOR MODEL VERIFICATION

An ideal system capable of detecting systematic errors in a numerical model would be one which could incorporate a number of techniques having enough utility to be applied to the least possible amount of data fields and yet provide a comprehensive examination of a model forecast. In Section 3 we described a number of methods which have this type of utility. We can now concentrate on the data which could be used. Based upon various studies of model performances at various weather centers (Miyakoda et.al., 1972; Arpe et.al., 1976; Fawcett, 1969) we can recommend that the following variables be used for verifying a model simulation.

Also based upon these same studies we can recommend that the variables and techniques can be applied to the following levels;

- i) 1000 mb;
- ii) 850 mb;
- iii) 700 mb;
- iv) 500 mb;
- v) 200 mb;
- vi) 100 mb;

The number of levels could be reduced at a later time.

4.1 Basic Analysis

The primary model parameter used in conjunction with the more basic techniques described in Section 3 would be the height field.

The height field can be analyzed for the above mentioned pressure surfaces. These levels provide information for the lower, middle and upper levels of the atmosphere, all of which play important roles in synoptic scale features. However analyses of height in the tropics would not be adequate due to the low variability there.

The height field is used with most of the basic methods described above. As an example, the application of

the Hovmöller diagram to the height field is useful for detecting trough and ridge positions.

The temperature field is also a useful tool for evaluating a model's performance. Temperature fields can be utilized, for instance, in latitude-height plots which can be particularly useful when examining the energetics of a model in a more detailed study.

Wind data is an important parameter to verify in a basic analysis. The parameter would be of interest to the FNOC product users due to the importance of wind in connection with wave conditions. Wind fields are also a better measure for evaluation in the tropics than height or temperature.

Another variable commonly used for model verification but which may be of less importance to FNOC users is precipitation. Variables such as precipitation, cloudiness, wind gusts, etc., which are produced by the model and may be used by a meteorologist as a basis for their own forecast can be a useful verification tool. These types of parameters are less important than height, wind or temperature, however they are easily verified using methods described in Section 3.

All of the variables mentioned here are available at the above mentioned levels. Also variables such as height, temperature and wind have great utility (e.g. deriving thickness from height or jet stream position from wind) and can be used to highlight many features of a numerical model.

4.2 Detailed Analysis

The parameters of height, temperature and wind are vital to a detailed analysis. Spectral techniques can be applied to all three variables. These variables are also necessary for the computation of quasi Lagrangian budgets and quasi geostrophic quantities.

Vertical velocity is also needed for a detailed analysis. Vertical velocity is available directly from model derived fields (omega) or can be calculated using other fields.

A comparison of the two velocity fields might show interesting features (Pearce, 1974).

Defining energy parameters as Lorenz (1967), requires the parameters of wind, temperature and height. Vertical velocity is needed to compute generation of dissipation terms.

The purpose of some detailed techniques being to use higher order parameters for diagnosing model performance can be accomplished with the four above mentioned data fields. Other techniques such as pattern recognition and regression analysis which do not necessarily require higher order terms can also use these variables.

4.3 Summary

The following variables can be used to verify a model forecast at the above mentioned levels;

- i) Height;
- ii) Wind;
- iii) Temperature;
- iv) Vertical velocity.

These variables are the basic atmospheric parameters relevant for synoptic scale processes. Many other parameters and features can be derived from these four variables. Therefore it is not necessary to save a lot of unneeded data which can be quite costly and wasteful. Also all of these parameters are readily available within the model system. However while height, temperature and wind are easily attained with a reasonable degree of accuracy, vertical velocity is not easily attainable in the actual atmosphere.

5.0

COST ESTIMATION

We can estimate the cost of implementing a procedure which would identify systematic errors in a numerical model by summarizing the impacts upon the present computer and software organization at FNOC plus the computational and plotting cost defined in terms of computer usage. The impacts made by a new system upon an existing system are often intangible in that it is very difficult to assign a representative dollar value on such items. The objectives of this section is to outline what impacts would be made through the implementation of such a system as described in Section 2. Any assignment of a dollar value on such impacts would not have much meaning.

Computer costs, in terms of time needed for specific calculations or the generation of certain plots are more tangible and easily linked to a specific dollar value.

5.1

Impact Summary

A summary of impacts made by the procedure described in Section 2 upon the existing operational system at FNOC can be divided into the following categories:

- i) Equipment impact;
- ii) Software impacts;
- iii) Organizational impacts;
- iv) Operational impacts;
- v) Developmental impacts.

5.1.1

Equipment impacts

The system described in this report would have a minimal impact upon the existing equipment at FNOC. The types of analyses and computations necessary for the system are straight forward and would not require specialized equipment or equipment configuration. However, the equipment used will determine such parameters as the timing of the system.

Equipment needed, besides a computer mainframe and data storage devices, are the VARIAN plotter and the Naval Environmental Data Network (NEDN). The NEDN component of the FNOC system would be used to provide the FNOC users with verification information necessary for the improvement of their products as described in Section 2.

5.1.2 Software impacts

The necessary software (programs and code), intended for the implementation of the system described in Section 2 will utilize existing software components (e.g. libraries, data bases) on the FNOC system.

The major software impact is in providing for the storage of data, needed for verification, which are not currently stored. This will require modifications to the software which chooses the data stored beyond the normal retention period within the FNOC data base.

5.1.3 Organizational Impacts

The system described in this report will require some organizational interaction. The basic analysis is designed to provide objective measures which can be reviewed by organizations with a vital need for specific information and/or a quality control organization at FNOC which could interact with users who have special problems or requests. This type of organizational impact is necessary for the improvement of the numerical product, which is the purpose of the system.

The detailed analysis is designed for interpretation and review by a central organization familiar with the physical details of the respective operational model.

Impacts upon the organizational system at FNOC would be in the form of assigning personnel certain responsibilities in terms of the interpretation of the verification measures and analyses. Added responsibilities may be placed upon specific organizations who must interpret results or interact with certain users more than currently practiced.

5.1.4 Operational Impacts

The operational environment at FNOC will be impacted when the verification system is run on a routine basis. As described in Section 2, it is recommended that the basic analysis be run routinely in conjunction with the operational model. As stressed above, the timeliness of the output for the basic analysis is an important component of the system. This impacts the operational environment in that the basic analysis should be scheduled to run as soon as the necessary numerical generated fields are available.

Another major impact is that the operational center will be required to interact with the user in terms of feedback regarding the possibility of a detailed analysis or the storage of different data.

The operational environment will be impacted by the necessity to store more data in a timely fashion. It may be that data needed to verify an important parameter is not stored routinely so the operational center would be required to provide for the storage of the necessary data.

A final impact is made by the restrictions that the output be produced under time constraints and made readily available for interpretation. This is an important component necessary for a successful operation. Specific printed and plotted output will be required in a timely fashion with respect to the operation of the numerical model.

5.1.5 Developmental Impacts

Prior to the implementation of the methods for determining systematic errors in a numerical model, impacts will be made upon other FNOC system users due to the use of equipment and data fields necessary for the testing and evaluation of the software. Necessary data will need to be stored and made available for testing. The development of software which would save data not presently archived will impact the operation of the altered data bases.

A limited impact upon the NEDN system will be made in order to develop the capabilities to distribute certain results to certain users.

5.1.6 Conclusion

The above discussions suggests that the greatest impact caused by the implementation of a system designed to identify systematic errors in a numerical model will be made upon certain organizations and the operational environment. The organizational impact is caused by the importance of the interpretation and user feedback components of the system. The operational impacts are primarily caused by the constraints that the analysis needs to be performed in a timely manner with respect to the operational model; more data may need to be stored than is saved under the present system and output may need to be distributed to specific users (e.g. Fleet Weather Centrals). These are important components of the system, necessary for a successful operation, however these impacts are quite intangible in terms of assigning a dollar value cost.

5.2 Computer Resources

We can make a rough estimate of the computer resources necessary for various types of calculations and plots by using estimates made from current existing software packages which perform similar analyses. All times would be representative of a CDC 6500 system.

5.2.1 Basic Analysis Resources

The basic analysis involves three main components:

- i) Data acquisition;
- ii) Analysis;
- iii) Output.

The data acquisition component will need to access ZRANDIO data bases for specific data fields. The amount of time necessary to accomplish this depends upon the structure of the data, the disk activity and the number of data bases needed. These factors will always vary, making an estimate of resources needed virtually impossible to make.

The analysis section will evaluate non-spectral verification measures and comparisons. These types of calculations and comparisons are straight forward and require a small amount of resources. Usually programs can be written which require small amounts of storage and no extended core storage (ECS). A similar type program which evaluates measures such as RMSE requires 2 cp seconds per comparison per FNOC data field.

The output section will be mainly dominated by the plotting requirements. Existing programs which can generate fields for Hovmöller plots using spectral data requires 14 cp seconds per latitude per level. In order to generate the actual plots a larger program requiring 110,000 octal words of central memory (CM) and 50,000 octal words of ECS is required. This routine requires 30 cp seconds to generate a plot which represents 10 days of data. The generation of these type of plots can be resource consuming, however they are not required to run very often. This would keep costs low. It is important that the system have the flexibility to limit unwanted more costly components.

Summarizing, we see that the input and output components of the basic analysis are the more costly components of the system. The analysis is straight forward and runs very fast on the computer. It would be possible to perform the basic measures within a few seconds of CP time. The more plots required, the more resources will be required for output. Difference fields are easily computed and contour plots using existing VARIAN plotter software require 30 cp seconds per field and 150,000 octal words of ECS as a minimum. Therefore difference and contour plots can use large amounts of resources.

The resource usage of the basic analysis is totally dependent upon input data used and the number of plotted outputs desired.

5.2.2 Detailed Analysis Resources

As discussed in the previous section the number of input fields and plotted displays required greatly dominate the amount of resources needed. However the analysis component for the more detailed investigation often requires complex calculations requiring more storage and time. For example we can estimate, based on existing routines needed to perform the calculations, that approximately 5-10 cp seconds would be required to convert the grid point data to a spectral representation using the standard 63 x 63 FNOC field. This would also require usage of ECS. An existing program which evaluates spectral evaluations of the RMSE, correlation and energy components requires 5 seconds per calculation and no ECS. We can estimate the resources necessary to plot contours of these quantities using the existing VARIAN software and similar existing programs. Each spectral calculation requiring a contour plot requires 150 cp seconds to execute and 47,000 octal words are needed to plot. Therefore a spectral energetics analysis would require approximately 150 seconds and 47,000 octal words of ECS per calculation, obtaining a 63 x 63 contoured field.

Quasi geostrophic analysis and regression analysis require less time than needed for a spectral analysis. The former two analyses involve straight forward numerical and statistical calculations. An evaluation of the quasi geostrophic equations would require less than half the time needed for a spectral analysis.

Quasi Lagrangian techniques are more complicated computationally than either quasi geostrophic or regression analysis. Therefore these type of calculations would require about the same amount of resources described in connection with the spectral analysis.

Pattern recognition techniques as described by Holl and Cuming (1979) for the MOSS system requires approximately 50 seconds per input field pair.

We can summarize by saying that techniques such as spectral analysis and energetic calculations and quasi Lagrangian techniques will require on the order of 100-200 cp seconds to evaluate parameters and make necessary plots. Other techniques such as pattern recognition, quasi geostrophic analysis and regression analysis generally require less than 100 seconds to operate.

5.3 Development Resources

Development resources pertain to the amount of time needed to develop, test and evaluate the software components of the system. Fortunately most techniques described in connection with model verification are widely used and many routines and "tools" exist which make development easier.

The development of the basic analysis would be straight forward. Most calculations are easily programmed and tested. More difficult tasks would involve the development of the input data and output plots, plus the interaction with the necessary system components needed to provide users with the necessary information. However once necessary input routines are developed they can be used for both the basic and detailed analyses.

Depending upon the availability of necessary data and equipment, a realistic estimate of the time needed to develop and test a basic analysis would be 1-2 man/months.

Input and plotting packages developed for the basic analysis could be applied to the detailed analyses. However the detailed analysis would be complicated to program. Fortunately many spectral routines exist and calculation of numerical equivalent of the quasi geostrophic quantities are straight forward. Also there are many regression packages available. The development, testing and evaluation of a detailed analysis system could be estimated at requiring 4-5 man/months to complete.

5.4

Cost Summary

We can summarize the cost estimation of a system for identifying systematic errors in numerical models as follows.

In terms of intangible cost measured as impacts upon existing procedures we suggest that this type of system would impact the operational and organizational components of the FNOC system most. Operational impacts would be caused by the inclusion of the system into the operational job stream and constraints upon the system for timely and special purpose runs and output. Impacts upon the organizations would be made by requiring personnel to be assigned for interpretation of results and interacting with users.

Resource costs are dependent upon the amount of data required to perform the analyses and the amount of plotted output required. The easiest way to keep resource costs low is to restrict the amount of costly plots to those needed only. The analysis portion for the basic investigation requires very little resources, however the detailed analysis requires much more resources and development time than the basic analysis.

Within this report we have described a plan which could be implemented to identify systematic errors in a numerical model. We have also described a number of techniques which can be used within the plan. The purpose of this study was not to say what techniques or plan design should be used. However we can draw a number of important conclusions which can be used as guidelines for a decision to implement methods to identify systematic errors in a numerical model.

The constraints of timely and concise output is essential for identifying errors in numerical forecasts for the purpose of improving the product. This is necessary for avoiding the generation of an unorganized set of statistics which can not be interpreted in any relevant fashion. It would be very easy to generate large set of measures for a large set of variables but only a few would actually indicate a possible error.

It is necessary to identify which model parameters will be analyzed to detect errors in numerical forecasts. This prevents saving and storing too much data which is very costly in terms of resource usage for storage space and accessing time for input into the analysis section.

The most important point to emphasize in the design of any system which would identify errors in numerical model forecasts is the importance of the interpretation of results. This is connected with the importance of feedback from the organization or organizations interpreting aspects of the verification results to the group performing the analysis and monitoring the development of the numerical model.

It is also very important to insure the statistical validity with respect to using various techniques and assigning an interpretation to the results. It has been emphasized within the discussion of the basic analysis and techniques applicable

to it, that many statistical methods are relatively simple to compute however more often than not steps are not taken which insure the validity of the results. This applies, for instance, to using small sample sizes and trying to attach a significant degree of reliability to results. Another application is the interpretation of global scores without accounting for regional biases. Before any specific statistical measures or techniques are applied to model verification one should determine whether the criteria for properly interpreting the results can be met. This can involve, for example, building a data base which spans a specific period of time.

It may also be necessary to determine whether certain errors in numerical models are truly systematic and not just smoothed fields of rapidly varying, random errors. It is necessary to determine the time scale of errors so that time periods used for analyzing whether systematic errors exist are truly representing systematic errors. For instance analyzing daily data may be inconclusive because errors are varying rapidly while a average of number of days may smooth the rapidly varying error components enabling the systematic error to be analyzed.

This report has suggested a generalized and flexible software plan which could be applied to identify systematic errors in numerical forecasts. The plan is applicable to different types of models and can be altered to operate in different modes. Therefore those organizations more familiar with certain errors or error patterns can govern the analysis more carefully.

A number of techniques applicable for use within the plan have been defined in terms of advantages and disadvantages.

An estimate of possible impacts upon the existing operational environment at FNOC plus the usage of resources needed by a system has been made.

This type of plan, for identifying errors, in a numerical forecast is best suited to provide the flexibility and interpretation needed to sufficiently analyze a forecast for the intended purpose of improving the FNOC product.

6.1 Summary

The following table is designed to summarize the techniques discussed in this report in terms of their utility in identifying systematic errors in the operational numerical model used at FNOC. The categories used to evaluate the techniques are described as follows;

- i) Existence - This is checked if there is a current software package at FNOC which implements this technique;
- ii) Interpretation - This is checked if the results of a particular technique can be interpreted in a quick, easy and straight forward manner. A check, followed by a minus (-) indicates that, for a particular technique (mainly graphical) minor post analysis is needed. An example is the need to evaluate a "t" statistic contour field with respect to a difference field which would indicate significantly varying areas;
- iii) Timeliness - This is checked if the results of the analysis can be obtained in a timely manner in an operational environment such as that existing at FNOC;
- iv) Complex Program - This is checked if the particular technique requires extensive and complex computer programs to operate;

- v) Cost - This is checked if the particular technique is costly in terms of computer resource usage;
- vi) Merit - This box is checked if the particular technique is widely applicable to a number of variables and applications for the identification of systematic errors in numerical forecasts.

Recommendations for the inclusion or exclusion of a particular technique, capable of identifying systematic errors in a numerical forecast model can only be made after all the above factors are evaluated with respect to a particular system or user need.

METHOD	EXISTENCE	INTERPRE- TATION	TIMELI- NESS	COMPLEX PROGRAM	COST	MERIT	RECOM- MENDA- TION
MSE	✓	✓	✓			✓	
CORRELA- TION	✓		✓				
BRIER SCORE		✓	✓				
DIFFER- ENCE FIELD	✓	✓	✓			✓	
HOVMOLLER DIAGRAM	✓	✓	✓			✓	
CROSS SECTIONS		✓	✓				
S1 SCORE	✓	✓	✓			✓	
CONTIN- GENCY TABLES		✓	✓			✓	

TABLE 6.1 Summary of the various techniques capable of identifying systematic errors in a forecast model.

METHOD	EXISTANCE	INTERPRE- TATION	TIMELI- NESS	COMPLEX PROGRAMS	COST	MERIT	RECOM- MENDA- TION
THREAT SCORE		✓	✓			✓	
SPECTRAL ANALYSIS	✓	✓-		✓		✓	
QUASI LAGRAN- GIAN				✓	✓		
QUASI GEOSTRO- PHIC				✓	✓		
REGRES- SION			✓			✓	
MOSS	✓	✓-	✓			✓	
PATTERN RECOGNI- TION		✓-	✓			✓	
DIAGNOS- TIC STUDIES		✓-		✓	✓		

Table 6.1 (continued)

REFERENCES

- Arpe, K., L. Bengtsson, A. Hollingsworth and Z. Janjic, 1976: A case study of a 10 day prediction. ECMWF Technical Report No. 1. 105 pp.
- Baumhefner, David, 1976: Present and future model verification methods applied to forecast results from the NCAR GCM. Weather Forecasts and Weather Forecasting: Models, Systems and Users. NCAR/C2-5+1976-ASP, 737-743. National Center for Atmospheric Research, Boulder, Colorado.
- Baumhefner, David, and Patrick Downey, 1978: Forecast Intercomparisons from Three Numerical Weather Prediction Models. Mon. Wea. Rev., Vol 106., 1245-1279.
- Bengtsson, Leonard, 1976: Verification activities at ECMWF in Scandinavia. Weather Forecasting and Weather Forecasts: Models, Systems and Users. NCAR/C2-5+1976-ASP., 721-733. National Center for Atmospheric Research, Boulder, Colorado.
- Brier, Glenn W., 1950: Verification of Forecasts Expressed in terms of Probability. Mon. Wea. Rev., Vol. 78, pp. 1-3.
- Brier, Glenn W., 1975: Statistical Questions relating to the Validation of Air Quality Simulation Models. U.S. Environmental Protection Agency, Office of Research and Development, National Research Center, Raleigh, N.C., 22 pp.
- Brier, Glenn W., 1976: Some comments on Model Verification and Forecast evaluation. Weather Forecasting and Weather Forecasts: Models, Systems and Users. NCAR/C2-5+1976-ASP, 704-709. National Center for Atmospheric Research, Boulder, Colorado.
- Brier, Glenn W., and Roger A. Allen, 1951: Verification of Weather Forecasts. Compendium of Meteorology, American Meteorological Society, Boston, Mass. 1334 pp.
- Daley, Roger, 1976: Verification Procedures at the Canadian Meteorological Center. Weather Forecasting and Weather Forecasts: Models, Systems and Users. NCAR/C2-5+1976-ASP, 710-720. National Center for Atmospheric Research, Boulder, Colorado.
- Fawcett, E.B., 1969: Systematic Errors in Operational Baroclinic Prognoses at the National Meteorological Center. Mon. Wea. Rev., Vol. 97, 670-682.

REFERENCES CONTINUED

- Freund, John E., 1972: Mathematical Statistics. Prentice-Hall Inc., Englewood Cliffs, New Jersey, 463 pp.
- Holl, Manfred M. and Michael J. Cuming, 1979: The Application of Measures of Synoptic Similarity to the Evaluation and Development of Atmospheric Prediction Models: Task 1. NAVENPREDRSCHFAC Contractor Report CR 79-01, 88 pp.
- Holton, James R., 1972: An Introduction to Dynamic Meteorology. Academic Press, New York. 319 pp.
- Houghton, David D. and William S. Irvine Jr, 1976: A Case Study Comparison of the Performance of Operational Prediction Models Used in the United States. Mon. Wea. Rev. Vol 104, 817-826.
- Johnson, Donald R, and W.K. Downey, 1975: Azimuthally Averaged Transport and Budget Equations for Storms: Quasi Lagrangian Diagnostics 1. Mon. Wea. Rev. Vol. 103, 967-979.
- Johnson, Donald R. and William K. Downey, 1978: The Mass, Absolute Angular Momentum and Kinetic Energy Budgets of Model Generated Extratropical Cyclones and Anticyclones. Mon. Wea. Rev. Vol. 106, 469-481.
- Lorenz, E.N., 1967: The Nature and Theory of the General Circulation of the Atmosphere. WMO, 161 pp.
- Miyakoda, K., G.D. Hembree, R.F. Stricler and I Shulman, 1972: Cumulative Results of Extended Forecast Experiments 1. Model Performance for Winter Cases. Mon. Wea. Rev., Vol. 100, 836-
- Payne, Stephen W., 1979: A Verification Case Study of the NEPRF/UCLA Global Forecast System. Fourth Conference on Numerical Weather Prediction, Silver Spring, Md. Oct. 29- Nov. 1, 1979.
- Pearce, R.P., 1974: The Design and Interpretation of Diagnostic Studies of Synoptic Scale Atmospheric Systems. Quart. Journ. Roy. Met. Soc. Vol. 100, 265-285.
- Somerville, Richard C.J., 1977: Pattern Recognition Techniques for Forecast Verification. Third Conference on Numerical Weather Prediction, Omaha, Nebr. April 26-28, 1977.

REFERENCES CONTINUED

- Trackton, M. Steven and J.D. Stackpole, 1976: The National Meteorological Center's Objectives, Programs and Plans for Verification and Evaluation of Model Performances. Weather Forecasting and Weather Forecasts: Models, Systems and Users. NCAR/C2-5+1976-ASP, 734-736. National Center for Atmospheric Research, Boulder, Colorado.
- Ward, J.H., P.J. Smith and D.P. Baumhefner, 1977: A Comparison of Observed and Predicted kinetic Energy Fields over North America. Third Conference on Numerical Weather Prediction, Omaha, Nebr., April 26-28, 1977.
- Wash, Carlyle H. and D.R. Johnson, 1977: A quasi-Lagrangian Budget Study of a PE Model Prediction of Cyclogenesis with "Locked-in-Error". Third Conference on Numerical Weather Prediction, Omaha, Nebraska, April 26-28, 1977.
- Wash, Carlyle H, 1979: Dynamical Verification of Numerically Simulated Extratropical Cyclogenesis using Quasi-Lagrangian Diagnostics. Fourth Conference on Numerical Weather Prediction. Silver Spring, Md. Oct. 29- Nov. 1, 1979.

Appendix A

The following is a brief summary of the basic characteristics of the models used in this report to illustrate verification techniques examined in the text. These models were studied in previously conducted verification studies (Arpe et.al., 1976; Baumhefner and Downey, 1978; Daley, 1976).

<u>MODEL</u>	<u>HORIZONTAL RESOLUTION</u>	<u>VERTICAL RESOLUTION</u>	<u>VERTICAL COORDINATE</u>
NCAR	2.5 degrees	4 layers	Z
GISS	4x5 degrees	7 layers	sigma
NMC	4 degrees	3 layers	sigma
UCLA	450 Km	6 layers	sigma
GFDL-N24	420 Km	9 layers	sigma
GFDL-N48	210 Km	9 layers	sigma

SPECTRAL MODELS

<u>MODEL</u>	<u>VERTICAL RESOLUTION</u>	<u>TRUNCATION</u>	<u>VERTICAL COORDINATE</u>
CMC	Variable (5 layers used in the case shown in this report)	Rhomboidal (variable truncation point)	sigma
GFDL	9 levels	Rhomboidal at 30 waves	sigma

AD-A098 848

CONTROL DATA CORP MONTEREY CA
SURVEY OF AVAILABLE SYSTEMS FOR
FEB 81 P A HARR, W D GROSCUP

F/G 12/1
IDENTIFYING SYSTEMATIC ERRORS I--ETC(U)
N00228-79-C-X554

UNCLASSIFIED

NEPRF-CR-81-01

NL

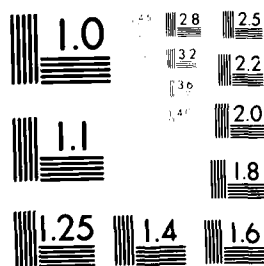
2 of 2
AD-A
[unclear]



END
DATE
FILMED
6-81
DTIC

OF 2

D A
0988848



MINIMUM RESOLUTION TEST CHART

Chief of Naval Research
Library Services (Code
Rm. 633, Ballston Tower
890 Quincy Street
Arlington, VA 22217

Chief of Naval Operatio
OP-952
Navy Department
Washington, DC 20350

Superintendent
U.S. Naval Observatory
ATTN: Dr. R. W. James
OPNAV 95201
34th A Massachusetts Av
Washington, DC 20390

Director
Attn: Library Acquisiti
20029, National Securit
Ft. Meade, MD 20755

Naval Deputy to the
Administrator, NOAA
Rm. 200, Page Bldg. #1
3300 Whitehaven St. NW
Washington, DC 20235

Officer in Charge
NAVOCEANCOMDET
AFGWC
Offutt AFB, NE 68113

Commanding Officer
Naval Research Lab
Attn: Library, Code 262
Washington, DC 20390

Commanding Officer
Office of Naval Research
Eastern/Central Regiona
Office, Bldg. 114, Ser
666 Summer St.
Boston, MA 02210

Commanding Officer
Office of Naval Research
1030 E. Green St.
Pasadena, CA 91101

Officer in Charge
NAVOCEANCOMDET
FLENUMOCEANCEN
Monterey, CA 93940

Commander
NAVOCEANCOM
NSTL Station
Bay St. Louis, MS 39529

Commanding Officer
NAVOCEANO, Navy Library
NSTL Station
Bay St. Louis, MS 39522

Commanding Officer
FLENUMOCEANCEN
Monterey, CA 93940

Commanding Officer
NAWESTOCEANCEN
Box 113
Pearl Harbor, HI 96860

Commanding Officer
NAVLASTOCEANCEN
McAdie Bldg. (U-117)
Naval Air Station
Norfolk, VA 23511

DISTRIBUTION LIST

Search Code 734) Tower #1	Commanding Officer NAVPOLARCOMCEN Navy Dept. 4301 Suitland Rd. Washington, DC 20390	Director Office of ENV. & Life Sci. Office of the Undersecretary of Defense for Research & Engineering (E&LS) Rm. 3B129, The Pentagon Washington, DC 20301
ations	Commanding Officer US NAVOCEANCOMCEN Box 12 COMNAVAMARIANAS FPO San Francisco 96630	Director Technical Information Defense Advanced Research Projects Agency 1400 Wilson Blvd. Arlington, VA 22209
ory es	Commanding Officer US NAVOCEANCOMCEN Box 31 FPO New York 09540	Director Systems Development NWS, NOAA Rm. 1216 - The Gramax Bldg. 8060 13th Street Silver Spring, MD 20910
ss Ave. NW 10	Superintendent Library Acquisitions U.S. Naval Academy Annapolis, MD 21402	Development Division National Meteorological Cen. NWS, NOAA World Weather Bldg. W32, Rm. 204 Washington, DC 20233
ositions urity Agency	Commander NAVAIRSYSCOM Attn: Library, AIR-00D4 Washington, DC 20361	Acquisitions Section IRDB-D823 Library & Info. Serv. Div. NOAA 6009 Executive Blvd. Rockville, MD 20852
AA #1 NW 35	Commander NAVAIRSYSCOM, AIR-370 Washington, DC 20361	Federal Coordinator for Meteor. Serv. & Sup. Rsch. 6010 Executive Blvd. Rockville, MD 20852
13	Commander NAVAIRSYSCOM Meteorological Systems Div., AIR-553 Washington, DC 20360	National Weather Service World Weather Bldg. Rm. 307 5200 Auth Road Camp Springs, MD 20023
2620 10	Commander NAVAIRSYSCOM, AIR-03 Washington, DC 20361	Head, Atmospheric Sci. Div. National Science Foundation 1800 G Street, NW Washington, DC 20550
Search ional l, Sect. D	Commander Pacific Missile Test Center Attn: Code 3250 Pt. Mugu, CA 93042	Director Div. of Atmos. Sciences National Science Foundation Rm. 644 1800 G Street, NW Washington, DC 20550
Search	Dept. of Meteorology Naval Postgraduate School Monterey, CA 93940	Laboratory for Atmos. Sci. NASA Goddard Space Flight Cen. Greenbelt, MD 20771
	Dept. of Oceanography Naval Postgraduate School Monterey, CA 93940	Executive Secretary Interdepartmental Committee for Atmos. Sciences National Science Foundation Rm. 312A 1900 G Street NW Washington, DC 20550
19529	Library Naval Postgraduate School Monterey, CA 93940	National Center for Atmos. Research Library Acquisitions P.O. Box 1470 Boulder, CO 80302
rary	Commander AWS/DN Scott AFB, IL 62225	Chairman Dept. of Meteorology Penn State University 503 Deike Bldg. University Park, PA 16802
19522	AFGWC/DAPL Offutt AFB, NE 68113	
	AFGL/LY Hanscom AFB, MA 01731	
	AFOSR/NC Boiling AFB Washington, DC 20312	
860	Commanding Officer U.S. Army Research Office Attn: Geophysics Div. P.O. Box 12211 Research Triangle Park, NC 27709	
)	Director Defense Tech. Information Center Cameron Station Alexandria, VA 22314	

ogy titute of	Nautilus Press, Inc. 1056 National Press Bldg. Washington, DC 20045
39	Oceanroutes, Inc. Attn: J. J. Maloney 3260 Hillview Ave. Palo Alto, CA 94304
ces Dept. cago et	The Executive Director American Meteorological Society 45 Beacon Street Boston, MA 02108
ysics ia at	American Met. Society Meteorological & Geostrophysical Abstracts P.O. Box 1736 Washington, DC 20013
0024	
ces Dept. nington	Australian Numerical Meteorology Research Center Attn: Dr. R. L. Hughes P.O. Box 5089A Melbourne Victoria, 3001 Australia
ces Dept. ersity 31	
ge of Science of Technology 19104	Library Atmospheric Environment Service 4905 Dufferin Street Downsview M3H 5T4, Ontario Canada
ogy ahoma	
ion	Meteorological Office Library London Road Bracknell, Berkshire RG 12 2SZ England
90406	
tment	European Centre for Medium Range Weather Forecasts Shinfield Park, Reading Berkshire RG29AX, England
. pee Rd.	
5440	

ATE
MED